DOCUMENT RESUME

ED 327 572 TM 016 006

TITLE	The Effects of Testing Project: The Effects of Testing on Teaching and Learning.
INSTITUTION	Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Nov 90
CONTRACT	OERI-G-86-0003
NOTE	71p.; Prepared by the UCLA Center for the Study of Evaluation in collaboration with the University of Colorado, NORC at the University of Chicago, and Arizona State University. The two papers were presented at the Annual Meetings of the California Educational Research Association (Santa Barbara, CA, November 1990) and the American Educational Research Association (Boston, MA, April 16-20, 1990), respectively.
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Conference Papers; Educational Improvement; *Elementary School Teachers; Elementary Secondary Education; Interviews; *Learning Processes; *Norm Referenced Tests; Psychometrics; Questionnaires; *Secondary School Teachers; *Standardized Tests; Teaching Methods; Test Results; *Test Use; Theory Practice Relationship
IDENTIFIERS	*Testing Effects

ABSTRACT

Two papers are presented as part of the Effects on Testing Project. The first paper, "The Effects of Testing on Teaching and Learning" (Joan Herman, Shari Golan, and Jeanne Dreyfus), describes a study focusing on standardized, norm-referenced tests. A questionnaire was administered to 85 teachers (kindergarten through grade 12 levels) attending a conference. The 131-item questionnaire was designed to determine: (1) the effects of mandated norm-referenced testing on curriculum and teaching; (2) variables that mediate these effects; and (3) the extent to which the results of testing represent school improvement. The study found that significant pressure on teachers improved test scores, teacher attention to test scores, and instructional time devoted to testing. Teachers did not report that emphasis on testing was narrowing the curriculum. The second paper, "Psychometricians' Beliefs about Learning" (Lorrie A. Shepard), examined, through interviews, the beliefs of 50 school district test specialists about learning. A majority operated from implicit learning theories encouraging close alignment of tests with curriculum and the judicious teaching of tested content. Beliefs associated with criterion-referenced testing were associated with a model of learning requiring sequential mastery of learning skills and behaviorally explicit testing of each learning step. Two tables and nine figures are appended to the second paper. (SLD)

ED322572

C·R·E·S·S·T

Center for Research on Evaluation, Standards, and Student Testing

TM

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. C. BEER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Center for Research on Evaluation, Standards, and Student Testing

Final Deliverable - November, 1990

The Effects of Testing Project

The Effects of Testing on Teaching and Learning

TM016006

BEST COPY AVAILABLE

TM

Center for Research on Evaluation,
Standards, and Student Testing

Final Deliverable - November, 1990

The Effects of Testing Project

The Effects of Testing on Teaching and Learning

Project Director: Joan Herman

Grant Number OERI-G-86-0003

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles

The first paper which follows was developed by Joan Herman, Shari Golan, and Jeanne Dreyfus. Ms. Dreyfus presented the paper at the annual meeting of the California Educational Research Association, Santa Barbara, California, November, 1990.

The second paper is another product of the Effects of Testing project, authored by Lorrie Shepard.

The research reported herein was conducted with partial support from the U.S. Department of Education, Office of Educational Research and Improvement, pursuant to Grant No. G0086-003. However, the opinions expressed do not necessarily reflect the position or policy of this agency and no official endorsement by the agency should be inferred.

The Effects of Testing on Teaching and Learning

Testing has assumed a prominent role in recent efforts to improve the quality of education. Viewing standardized tests as a significant, positive and cost-effective reform tool, educational policymakers have been using them at an increasing rate. The testing process now costs hundreds of millions of dollars and thousands of hours of administrative, teacher and student time.

The reasons for the increased use of testing are many. Following advice from testing advocates, policymakers believe that testing sets meaningful standards to which school systems, schools, teachers, and students can aspire; that test data can help shape instruction; that it serves important accountability purposes; and that coupled with effective incentives and/or sanctions, testing is a powerful engine of change. As evidence of the latter, proponents point with pride to rising test scores.

Yet while testing is thought by many to benefit education in a variety of ways, and recent policy anoints it as a major carrier of reform and change, the validity and value of traditional standardized forms of testing are subjects of increasing debate. Recent studies raise questions about whether improvements in test score performance actually signal improvement in learning (Cannell, 1987; Linn, Grave and Sanders, 1989; Shepard, 1989). Other critics take issue with the narrowness of content of such tests, their match with curriculum and instruction, their

neglect of higher level thinking skills, and the relevance and meaningfulness of their multiple choice formats (Baker, 1989; Shepard, 1989, Herman, 1989). According to these and others, rather than exerting a positive influence on students' learning, testing has trivialized the learning and instruction process, has distorted the curriculum, and usurped valuable instructional time for some students. (Smith, Edelsky, Draper, Rotytenber, and Cherland; Romberg, Zarinnia, and Williams, 1989; Bracey, 1989; Stake, 1988; Dorr-Bremme and Herman, 1986)

Testing, thus, has produced important yet debatable changes in our educational system and numerous studies have looked at some of these changes in depth. Those that are pertinent to this study are reviewed below.

New Driving Frameworks

Changes in the educational environment in the last twenty years have reshaped the conceptual frameworks and major themes that researchers consider when they study testing and its effects. Increased government funding to schools and growing public concern about the quality of education in the U. S. have raised the level of accountability for all involved - teachers, administrators and state educational personnel. This increased accountability has had two major effects. It has increased the "stakes" or the consequences of testing and it has also fostered the concept of measurement-driven instruction.

Testing in many states and schools districts is now a "high" stakes process. Testing is defined as high stakes when test results are thought to influence important decisions which state and local administrators make about such things as curriculum, program appropriations, student promotion, and teacher evaluation (Popham, 1987; Madaus, 1987; Romberg, Zarrinnia, Williams, 1989). The push for educational equality and excellence, increased federal financial aid to schools, and a greater public sentiment for accountability have all contributed greatly to raising the stakes of testing.

"High stakes" testing also reveals a new view of the role of measurement and testing in instruction. In the past, tests were not expected to affect curriculum or alter instruction. They served as a general barometer of educational quality. Today, though, the value of linking teaching to measurement - measurement-driven instruction (MDI) - is a hot topic. (Bracey, 1987; Popham, 1985, 1987) Testing itself is viewed as a reform and policy intervention. Those who embrace it argue that not only is it a cost-effective way to improve instruction, but it is needed to bring order to the haphazard situation that exists because of the proliferation of high-stakes testing that exerts significant influence on classroom learning. (Popham, 1987)

Critics of MDI say that it reverses the "normal order of things" and trivializes learning. (Bracey, 1989, pp. 684-685) Because measurement-driven instruction addresses specific

instructional objectives that can be easily assessed, opponents also believe that it fragments learning and may miss significant learning outcomes. According to Richard Richardson, a University of Arizona professor, and his colleagues MDI objectives promote "bitting" - little bits of information are parcelled out to students because that is what the MDI tests measure.

These same critics also believe that MDI deflects or shifts the focus of instruction to those things which are easily assessed, rather than significant knowledge acquisition and development of high level skills. They further believe that this shift trivializes the objectives that are tested, translating learning goals into multiple choice test questions. Higher order learning skills, in short, are given short shift. (Richardson, pp. 43-49)

Time on Testing

Dorr-Bremme and Herman (1986) found that for elementary school children "testing across the curriculum consumed eight to ten percent of students' available curriculum time." (Dorr-Bremme and Herman, p. 23). This study looked at all types of testing, from state and district mandated tests to teachers' classroom tests. Smith et al., in her study of two "high stakes" elementary schools (1989), found "somewhere between three and four weeks of school time" was spent on testing, and test preparation. (Smith et al., p. 267) This

did not include the time teachers and students spent on internal, teacher prepared tests.

The nearer in time to the test, the more time spent on direct test preparation. Twenty-eight percent of the teachers in Smith et al.'s study (1987) started two or more months before the test and an additional twenty-two percent started the week before. Ninety percent of the teachers in the study we involved in test-taking practice during the test week itself. (Smith et al., 1989, p. 284.)

Time spent on testing also appears affected by the number and type of tests given. In their study of the effects of mandated testing on math instruction, Romberg, et al. (1989), found that California teachers allocated instructional time according to which mandated test they had to administer. In their case, more time was spent on preparing for district tests than for the CAP test (California State Assessment Program). The teachers in their study also used the district test information much more than CAP information. They used district test results to group students and assign them to special programs, inform parents, and gauge themselves and their instructional program. (Romberg, et al., 1989, pp. 86-87, Appendix L).

How Testing Affects the Schools

Beyond impacts on instructional time, several researchers have examined how testing affects the school by looking at how it affects those involved - including

administrators, teachers and students. In addition, they have examined how testing affects classroom organization, curriculum decisions, teacher evaluation, and the over-all learning environment. In their national study of elementary and secondary school teachers, Dorr-Bremme and Herman (1986) found that in eight major school decisions or tasks (e.g., curriculum, student promotion, teacher evaluation), teachers' classroom testing provided more important information than any other types of test. They also found that "teachers' opinions, judgments,and recommendations clearly carry more weight than any type of test results." (Dorr-Bremme and Herman, pp. 32-33) Yet, studies done more recently point to a change in the effects of testing - especially in decisions concerning curriculum and instruction. (Smith, et al., 1987; Corbett & Wilson, 1988; Shepard, 1989)

Depending on your viewpoint, standardized testing coupled with increasing accountability pressures has prompted either an interest in or a concern about the linking of test content with curriculum taught. (Popham, 1985, 1987; Richardson, 1985; Bracey, 1987) The evidence regarding how often and to what extent this occurs is inconclusive. In their review, MacRury, Nagy and Traub (1987) found that there was little or no impact on curriculum with the introduction of large-scale assessment programs. (p. 13) Similarly, in their study on the influence of mandated testing on mathematics instruction (1989), Romberg, Zarinnia and Williams also found that the majority of the five hundred and

fifty-two teachers involved "do not increase or decrease their instructional emphasis because of the test nor do they consider the style and format of test items when planning their own instruction." (Romberg, et al., 1989, p. 33). This finding is also supported by the work of Ruddell (1985) in seven California districts. Sixty-one percent of the teachers involved in the study stated that standardized tests had little effect on what they taught.

Other studies, though, have yielded data which support the belief that standardized testing has influenced curriculum. Madaus (1988) found that if teachers believed that important decisions were tied to test scores, the teachers will teach to the test. The work of Smith and her colleagues (1987) supports this conclusion and examines in detail how curriculum is affected. Smith et al., (1987) found in the elementary schools they studied that "in high stakes environments, schools neglect material that the external tests do not include...reading real books, writing in authentic contexts, solving higher-order problems, creative and divergent thinking projects, longer-term integrative unit projects, computer education and such are gradually squeezed out of ordinary instruction." (Smith et al., p. 268) They cited science as an example of a nontested subject whose teaching had been negatively affected by the pressure to cover tested materials. They found that science, for example, "at the intermediate grades looks more like reading all the time." (Smith et al, p. 268) Teachers felt

that setting up science activities took too much time and as testing neared, the subject was dropped entirely to make way for test preparation. The elementary school teachers in Dorr-Bremme and Herman's study (62%) also believed that minimum competency requirements either already had or would adversely affect the amount of time spent on teaching subjects not included in the tests.

In their study of a high-stakes environment of mandatory minimum competency testing in Maryland and Pennsylvania, Corbett and Wilson (1988) had similar results. Curriculum was significantly impacted. Maryland schools, for example, in their attempt to improve scores, altered the curriculum, "especially in terms of redefining course objectives and resequencing course content". (Corbett and Wilson, 1988, p. 30)

Standardized testing is also affecting instructional techniques. In their desire to give adequate test preparation, teachers train the students in testing formats. Smith at al., (1987) found that teachers were using worksheets that duplicated the question layout of a standardized test. Teachers in their study used math drills and frequently administered timed tests. Spelling was taught and tested in a format similar to that which appeared on mandated tests. (Smith, et al., 1987)

In addition to studying the effects testing has on curriculum, many studies have examined the effects that testing has on staff. Mandated testing creates tension.

Corbett and Wilson (1988) found that "Maryland teachers were reported to be under greater stress...and to have experienced decreased reliance on their professional judgments than teachers in Pennsylvania." (where there was not a direct attempt to raise scores.) (Corbett and Wilson, p. 30) In her study of test score gains (1989), Shephard found that those involved in education had heard that dismissal of principals and/or superintendents had been tied to test results. In fact, this seldom happened, but the belief that it did caused anxiety for principals and staff.

Those studies that looked at student changes found that testing could have both over-all and specific negative effects on students. Primary grade teachers in Smith's et al. study felt that "tests injure the pupils' psychological well-being and sense of themselves as competent learners." (Smith et al., 1987, p. 217). They also cited a whole litany of negative effects during test week. For example, the teachers saw a rise in student truancy, stomach symptoms, worry, vomiting, crying, wetting, headaches and refusal to take the tests. (Smith et al., 1987, p. 284)

There are indications that testing impact may be highly related to scoioeconomic statistics. Dorr-Bremme and Herman found that, compared to high SES schools, administrators in lower SES schools were more influenced by formal tests results - "especially minimum competency measures and district objectives-based tests" - when making key decisions such as curriculum planning, funding allocations and

reporting test results to the public. (Dorr-Bremme and Herman, 1986, p. 34)

Testing Practices and Sources of Pressure

Administrators - both district and school-site - play a pivotal role in shaping the school testing environment. They can take a "top-down" approach and dictate what the curricula should be, and how the teachers should prepare the students for the test. On the other hand, they can provide some degree of guidance, in-service and resource materials but let the teachers shape the curriculum and decide what type of test preparation is best for the students. (Glickman, 1987) Whichever course they choose, their influence is apparent. Eighty percent of the teachers in Smith et al.'s study (1987) said that they "were encouraged (by administrators) to raise test scores." (Smith et al., 1987, p. 283) Seventy-five said that principals and district administrators also wanted them to teach test-taking skills.

In Shepard's study (1989) on test score gains, state testing directors reported that "presentation of test results to the state board is a media event" and that this coverage was the "most pervasive source of high-stakes pressure." (Shepard, p. 7; Corbett, Wilson, 1988) Where there is press coverage of test results, there is also editorializing. The pros and cons of the educational system are discussed in the public forum.

Many administrators agree that the public has a right to know about the status of educational achievement. In 1979, Michigan's educational directors made changes to its statewide testing program based on several "need to know" concepts. Among them were that the public has a right to know about the achievement levels of students in public schools and that they should be informed about the level of remediation when achievement scores are low. (Roeber, Donovan, Cole, 1980). In addition, they firmly believe that the news should come from the educational system and that results should not be "discovered" by the press.

Yet, this public pressure can have adverse effects. For a few, teaching to the test has turned into teaching the actual test and some districts have had to cope with outright cheating. In 1974 in New York City, for example, all schools were ranked on the basis of reading scores. Buckling under this pressure a few New York schools obtained the mandated test and used it to prepare students prior to the testing date. The "allegation was made that students, teachers and parents" were all aware. (Polemeni, 1977, p. 34)

In a March 13, 1990 Wall Street Journal article on toughening school testing, Arnold Fege, a lobbyist for the National Parent-Teacher Association, expressed educators' fear about testing. "What we're scared of is that we're going to do so much testing and so much assessing, we aren't going to have time to do any learning." (Putka, p. B1)

The study which follows seeks to clarify the debate about the effects of testing. It focuses on standardized, norm-referenced tests. The study employs an extensive teacher questionnaire and uses the data to assess the impact of these tests in several areas.

Methodology

The questionnaire study which follows was designed to answer the following questions:

1. What are the effects of mandated, norm-referenced testing on curriculum and teaching?
   - Does it influence what is taught?
   - Does it influence how it is taught?
   - What is the nature of test preparation?
2. What variables mediate these effects?
   - Teacher background and attitudes
   - School action
   - Pressure to improve test scores
3. To what extent do the results of testing represent school improvement?
   - To what extent do they represent changes in demographics?
   - How do educators perceive the reasons for the change - or lack thereof?

Subjects

The subjects were 85 kindergarten through twelfth grade teachers from a large urban school district who voluntarily chose to answer the questionnaire. They were part of a larger group attending a teacher leadership institute where the questionnaire was distributed. Fifty-five respondents were from elementary schools and thirty were from secondary schools. The teachers at both levels were experienced with an average of seventeen years in the classroom and eight years in their current school. Thirty-five subjects taught classes which had 0 to 25% Chapter I students, while 42 of them had 76% or more Chapter I students in their classes. (see Table 1 for details) A serious caveat of this study is that it is based on a small sample which may not be representative of the larger population of public school teachers.

Questionnaire

A teacher questionnaire containing 131 items was developed by the authors for this study. The questionnaire has four components with several sub-sections. The first component asks about teacher and student background and the school context in which testing takes place. The second part is concerned with test-taking strategies and test preparation practices. It inquires about the degree of focus on test content and test-taking skills and looks at staff development activities for test preparation. Component three deals with

testing's impact on instructional objectives, content taught, staff professionalism, and the degree of interference with sound instructional practices. The last questionnaire component looks at teachers' attitudes about testing, particularly their perception of why scores increase or decrease, of the controllability and stability of test scores and of the validity of test scores as a sign of academic achievement and school improvement.

Questionnaire results were analyzed by school level (elementary, secondary) and by the SES levels of the students served. For the purposes of these analyses, low SES was defined as those with at least 80% Chapter One students; high SES was defined as less than 20% Chapter One. Thus, in the analyses which follow, low SES and high SES do not constitute the entire sample. The whole sample, including the middle group, is captured in the "overall" means.

Findings

This study focuses on several important questions about the effects of testing. What are the actual effects of testing on curriculum and instruction? Who or what mediates the effect and to what extent? How much attention do school administrators and teachers pay to the testing process and test scores? What changes in instructional practices and activities, job climate and causes of test score movements have occurred over the last three years? And, what are teachers' attitudes toward testing and how are they affected

by the pressure to increase their students' test scores? The findings of this study supply some answers to these questions.

1. To What Extent Do Teachers Feel Pressure to Improve Test Scores?

Overall, teachers feel that the media, district school boards and administrators and principals exert the most pressure on them to improve test scores. Teachers serving low SES students report stronger pressure from these groups than do those serving higher SES students. Parents and the community were viewed as low sources of pressure for improvement. (see Table 2 for details)

2. How Much Attention Do Schools Give to Test Scores?

In general, elementary schools pay more attention to test scores than secondary schools do and their administrations engage in repeated activities with their teachers to review, monitor and improve test scores. Specifically, low SES elementary schools give the most attention to test results. In these schools, there are noticeably more, though infrequent, comparisons of teachers based on their students' test performance, and administrators (more than a few times) discuss with their teachers ways to improve scores and strengthen instruction in weak areas.

Typically, low SES elementary schools also provide teachers with practice test-taking materials more than once over the course of the year. Both secondary and elementary schools seldom consider test scores when evaluating teachers. (see Table 3 for details)

3. How Does School Attention to Test Scores Compare to Attention to Other Important Educational Issues such as New Instructional Ideas, Higher Order Thinking Skills and Student Attitudes Toward Learning?

Table 4 shows that the attention is roughly comparable. Note the repeated and relatively more frequent attention to higher order thinking and new instructional ideas in the low SES elementary group compared to other respondents. (See Table 4 for details.)

4. What is the Influence of Testing on Teachers' Instructional Planning?

To some extent, elementary school teachers, whether serving high or low SES students, review the test's objectives and the content and skills covered in the tests; look at old or current test to make sure their curriculum includes the test's content; and adjust their instructional plans based on their current students' most recent scores. While secondary schools pay somewhat less attention to test

results in their planning, we see strong differences between high and low SES at this level. Secondary teachers serving disadvantaged students show patterns generally similar to elementary school teachers. (see Table 5 for details)

5. How Much Class Time do Teachers Spend on Test Preparation?

In elementary schools, teachers spend the equivalent of several weeks in instructing students on test-taking strategies; give students about a week's worth of practice with test-item formats, and engage them in worksheets which review test content for several days to a week. Secondary teachers spend slightly less time on each type of preparation. Elementary teachers and secondary teachers serving low SES students report spending more time overall on test preparation than do secondary teachers serving higher SES students. Teachers on both levels seldom give students old forms of the test on which to practice, but do generally use commercially developed practice materials. (see Table 6 for details)

6. What are Teachers Attitudes about Testing?

Expectations. Both elementary and secondary teachers have moderate to strong expectations that their students will do well on their standardized test. Secondary teachers

teaching low SES students are the most positive on this dimension and as shown by the standard deviation, the most "consistent" (i.e., in agreement). On other indicators, teachers at both levels tended to modestly agree that they could influence their students' test scores. (see Table 7 for details)

Pride. All groups felt that teachers at their schools have a strong sense of pride in their work, particularly those serving higher SES students. And all groups tended to moderately disagree with the idea that schools were more interested in improving test scores rather than overall student learning. (see Table 7 for details)

Helpfulness. Overall, elementary school teachers, especially those serving low SES students, do not believe that testing is helping schools improve or clarify important learning goals, nor do they feel that it gives important feedback. Secondary teachers show similar, though slightly less pessimistic, views. While almost all feel that testing creates tension for them and their students (there were only a few negative responses to this item), the elementary school sample expressed stronger and more universally negative feelings. (see Table 7 for details)

Fairness. None of our subjects perceived the tests as particularly fair. While all groups were somewhat neutral to slightly positive about whether they can substantially influence how well their students do, they do not generally believe that changes in test scores are reflective of their

teaching. Furthermore, teachers at all levels were consistent in the belief that there is a discrepancy between what should be taught and what the test emphasizes. (see Table 7 for details)

The next set of questions and analyses examine differences in responses depending on whether teachers teach in schools where test scores are going up, declining, remaining the same, or fluctuating. To get a sense of the extent to which these score trends are confounded with SES and school level, table 8 shows the distribution. Here we see that teachers reporting increasing scores are relatively more likely to be low SES elementary schools while in our sample teachers reporting decreasing scores were relatively more likely to be in high SES elementary or secondary schools.

7. What Do Teachers Perceive as the Causes of Test Score Changes by Test Score Trends Over the Last Three Years?

Table 9 shows that teachers whose students' test scores have decreased or fluctuated over the last three years believe the cause to be more than moderately related to changes in student population, in school climate and in the community. Teachers whose students' scores have increased over the last three years, in contrast, believe that changes in teaching effectiveness have been a moderate factor (i.e., if scores get worse, it's due to changes in the environment;

if they get better, it's because their teaching is more effective). And, no matter what the status of test score changes, change in test administration practices was the least influential factor for all. Other conclusions are difficult to draw since the average ratings for the other factors were in a tight range from about 2.4 to 2.9. (see Table 9 for details)

8. How is Pressure to Improve Test Scores Related to Test Score Trends?

Teachers whose students' scores are decreasing feel greater pressure from a multitude of sources than do other teachers in our sample. (see Table 10 for details)

9. How is School Attention to Test Scores Related to Test Score Trends?

Schools in all test score trend groups report more frequent attention to basic skills instruction than to higher order thinking skills, particularly those in schools where scores are fluctuating or remaining the same. It is interesting to note that attention to these two areas is closest in schools where scores have shown an increase. (see Table 11) No clear differences in test score trend groups emerged in other indicators of school attention to testing.

10. How is Time Teachers Spend on Test Preparation Related to Test Score Trends?

Teachers with decreasing student test scores engage more often in various types of test preparation activities than any other test score trend group. In particular, they spend the most time, equivalent to almost a month, teaching test-taking strategies and a few weeks giving practice in the different test item formats. They also spend time giving students worksheets that review expected test content and, for at least a few days, use commercially produced practice tests with their students. These same teachers spend little time, about a day, giving students old test forms on which to practice. (see Table 13 for details)

11. How is the Extent of Instructional Renewal in Schools Related to Test Score Trends?

Instructional renewal is greater in schools with increasing scores than it is in schools with decreasing scores. In addition, for improving schools many aspects of this renewal have increased over the last three years, while for declining schools instructional renewal activities have remained the same. Teachers in our study whose scores were increasing, for example, see at least moderate attention to student interest in learning, stronger and increasing support for school wide or grade level planning, greater and

increasing programmatic efforts to improve student learning and more implementation of innovative instructional strategies than do teachers working in decreasing score schools. (see Table 14 for details)

12. How Is Attention to Other Academic Subjects Related to Score Trends?

With the exception of teachers whose test scores are increasing, all of the study's participants spend "a lot" of time drilling students in basic skills and give at least moderate attention to higher order thinking skills. The pattern for attention to both basic skills and higher order thinking skills has remained the same over the last three years.

Overall, teachers in our study said that subjects which are not included in the test receive moderate attention. Differences do exist by score trend in the amount of attention given to science. Those with decreasing or fluctuating scores give the most attention to science, while those with constant scores give the least. (Teachers with increasing scores fell in the middle but indicated that the amount of attention given to science has increased over the last three years.) Finally, teachers whose scores are decreasing clearly give the most time to test preparation. (see Table 15 for details)

13. How is Degree of Teacher Job Satisfaction Related to Score Trends?

Overall, teachers with decreasing student scores have the least amount of job satisfaction. This group believes that their ability to meet individual student needs has decreased over the first three years and of all score trend groups, the image of teacher as efficient educator is the least apparent in their schools. Yet, across the board, they and their peers in this study perceived that teachers' influence on school decision-making has increased over the last three years and, overall, they see themselves as have strong control over their classroom programs. (See Table 16 for details)

14. What Significant Correlations Exist Among School Characteristics, Teacher Attitudes, and Testing Variables?

We found that there are several significant correlations ($p=.05$) between overall pressure, overall time spent on test preparation, the number of Chapter I students and the effects of testing.

Pressure. Our data indicate that overall pressure to improve test scores has a positive correlation with overall school attention to test scores. It also is correlated with testing's overall influence on instructional planning and with overall time spent on test preparation. There is also a

negative correlation between overall pressure and teachers' perceived control over their classroom instructional program and their overall pride in teaching. (see Table 17 for details)

Planning Influence. Testing's influence on planning has a positive correlation with overall time spent in est preparation and the pressure to cove all required curriculum. It has a negative correlation with teachers' perceived control over their classroom instructional program. (see Table 17 for details)

Chapter I students. The number of Chapter I students and the effects of testing also are related. There is a positive correlation between the number of Chapter I students and overall pressure to raise test scores. The number of Chapter I students is also correlated positively with school attention to test scores, overall time spent on test preparation and pressure to cover all required curriculum. Conversely, there are negative correlations between the number of Chapter I students and overall pride in teaching and overall job satisfaction. (see Table 17 for details)

Conclusion

The purpose of this exploratory study was to examine the impact of standardized, nationally normed tests on curriculum and instruction and to ascertain what variables mediate the impact. Given the sample, our conclusions necessarily are very tentative. The study finds significant pressure on

teachers to improve test scores and significant school and teacher attention and instructional time devoted to testing. Certainly not surprising. However, one interesting finding is that the teachers did not report that emphasis on testing is narrowing their curriculum, as indicated by the attention they give to higher level thinking skills, subjects not tested, etc. There is some evidence, though, that testing is interfering with teachers' ability to attend to the finer details of instruction, i.e. attention to individual students, use of innovative instructional strategies and opportunities for student choice in what to study. Furthermore, given the sheer time and attention to testing, one wonders whether something necessarily gets short changed.

Our data suggest that teachers perceive themselves as giving some attention to everything, i.e., preparing students for the standardized test as well as teaching the required curriculum, the fine arts, science, and other subjects not tested. They also feel that they teach both basic skills and higher order thinking skills. And they indicated that although they do drill, they also engage their students in project and small group work. If this is representative of today's trend, the question is how long can teachers keep up this pace? Furthermore, when the next reform appears, how will they incorporate it into their already full teaching load and continue spending significant time and attention on testing without displacing something else? The implications

especially disadvantaged students, need to be given greater attention.

Finally, the study finds no clear relationship between reported test score trends and time and attention to testing. While there was some indicating of lower morale in schools with decreasing scores, it is interesting to note the positive climate and innovation in those with reported increasing scores.

The findings reported here are the result of a pilot study. The issues it raises will be more fully explored with a controlled and representative sample of teachers.

References

Bracey, G.E., (1987). Measurement driven instruction: Catching phrase, dangerous practice," Phi Delta Kappan, May.

Cannell, J.J. (1987). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. Second edition. Daniels, West Virginia: Friends for Education.

Corbett, H, D., and B. Wilson, (1988). Raising the stakes in statewide mandatory minimum competency testing," Politics of Education Association Yearbook. pp. 27-39.

Dorr-Bremme, D. W., & Herman, J. (1986). Assessing student achievement: A profile of classroom practices. Center for the Study of Evaluation, University of California, Los Angeles.

Glickman, C. D., (1987). Unlocking school reform: Uncertainty as a condition of professionalism, Phi Delta Kappan. 69(2), pp. 120-122.

Linn, R., Grave, & Sanders, (1989). Quality of Standard Tests: Final Report. Center for the Study of Evaluation, University of California, Los Angeles.

MacRury, K., Nagy, P., and R. E. Traub, (1987). Reflections on large-scale assessments of study achievement. The Ontario Institute for Studies in Education, Toronto.

Polemeni, A. J., (1976). Accountability and test security, The Journal of Teaching and Learning. 2(2), pp. 1-6.

Polemeni, A. J., (1977). Security in a citywide testing program, The Journal of Teaching and Learning. 2(3), pp. 34-40.

Popham, W. J., (1990). Appropriateness of teachers' test-preparation practices, presented at a Forum for Dialogue Between Educational Policymakers and Educational Researchers. University of California, Los Angeles.

Popham, W. J., (1983). "Measurement as an Instructional Catalyst," Ekstrom, R. B. (Ed.), Measurement, Technology, and Individuality in Education. 17, March, 19-30.

Popham, W. J., (1987). "The Merits of Measurement-Driven Instruction," Phi Delta Kappan. May, pp. 679-682.

Popham, W. J., Cruse, K. L., Ranking, S.C., Dandifer, P. W. and Williams, P. L. (1985). Measurement-driven instruction: It's on the road, Phi Delta Kappan. 66(9), pp. 628-634.

Putka, G. (1990). Educators decry U.S. push to toughen school testing, The Wall Street Journal. Dow Jones & Company, March 13.

Richardson, R. C., (1985). How are students learning?, Change. May/June, pp. 43-49.

Roeber, E. D., Donovan, D. L., and R. T. Cole, (1980). Telling the statewide testing story . . . and living to tell it again!", Phi Delta Kappan. 62(4), pp. 273-274.

Romberg, T. A., Zarinnia, E., and S. R. Williams, (1989). The influence of mandated testing on mathematics instruction: grade 8 teachers' perceptions. National Center for Research in Mathematical Science Education, University of Wisconsin-Madison.

Ruddell, R. B., (1985). Knowledge and attitude toward testing: Field Educators and Legislators," The Reading Teacher. pp. 538-543.

Shepard, L. A., (1989). "Inflated Test Score Gains: Is it Old Norms or Teaching the Test?" Center for the Study of Evaluation, University of California, Los Angeles.

Smith, K. F., (1987). "Effects of Testing on Teaching and Learning," James Cook University of North Queensland.

Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C., and Cherland M., (1987). "The Role of Testing in Elementary Schools," Center for Research on Educational Standards, and Student Tests and the Office of Educational Research and Improvement, Department of Education, University of California Los Angeles, and Arizona State University.

Stake, R. E., (1988). "The Effect of Reforms in Assessment in the USA," paper presented at the British Educational Research Association at the University of East Anglia.

Stake, R. E., (1989) "Quality Control and Deceptive Packaging: Does Reform Require Accourate Portrayal of Education?", paper presented for a public presentation at the University of North Carolina.

PSYCHOMETRICIANS' BELIEFS ABOUT LEARNING

CSE Technical Report 318

Lorrie A. Shepard

University of Colorado at Boulder

UCLA Center for Research on Evaluation, Standards, and Student Testing

April, 1990

The research reported herein was conducted with partial support from the U.S. Department of Education, Office of Educational Research and Improvement, pursuant to Grant No. G0086-003. However, the opinions expressed do not necessarily reflect the position or policy of this agency and no official endorsement by this agency should be inferred.

This paper was written as part of a research project sponsored by the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST). It was prepared for presentation at the Annual Meeting of the American Educational Research Association, Boston, April 17, 1990.

Please address inquiries to: CSE Dissemination Office, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, California, 90024-1521

In this paper I propose to examine the beliefs that psychometricians hold about learning. What models or conceptions of teaching and learning do measurement specialists invoke when they make mental decisions about testing practice? In proposing this line of inquiry, I am borrowing methodological approach and perspective from recent research on teacher thinking, which suggests that teachers' classroom practices can be understood in terms of their beliefs or implicit theories about instruction and learning. As described by Clark (1988), "These theories are not neat and complete reproductions of the educational psychology found in textbooks or lecture notes. Rather, teachers' implicit theories tend to be eclectic aggregations of cause-effect propositions from many sources, rules of thumb, generalizations drawn from personal experience, beliefs, values, biases, and prejudices" (p. 6). Similarly, psychometricians very likely have shared and idiosyncratic ideas about student learning and the role of testing in effective instruction.

The possibility that psychometricians and measurement specialists have unstated learning theories that influence their practices of testing and assessment was suggested by several observations. For example, in telephone interview data from 50 state directors of testing, there was almost uniform agreement among the 40 directors, who characterized their testing programs as having "high stakes," that high pressure tests focused more instructional time and attention on tested objectives (Shepard, 1990). However, respondents differed as to whether they attached a positive or negative "valence" to the teaching changes they perceived to occur in response to testing. By implication some believed that students in their state would learn more because high-stakes testing forced attention to important skills that had hitherto been neglected. In contrast, those who worried about the effects of testing on instruction believed that somehow something would be lost if the tests reshaped curriculum. These two groups did not appear to differ by the amount of reported pressure associated with testing nor by the type of test administered (i.e., norm-referenced or criterion-referenced), thus making more plausible the inference that differences in belief systems accounted for differences in respondents' interpretations of effects.

A similar difference in perspective can be seen in arguments about what constitutes legitimate test preparation. Mehrens and Kaminski (1989) conducted a content analysis of one version of the test preparation materials called *Scoring High* and found them to be so similar to the actual test that, in the judgment of Mehrens and Kaminski, using these materials would be the same as practicing with the test beforehand and therefore unethical. Makers of *Scoring High*, however, recommend that their materials be used daily for 4-5 weeks before regularly scheduled standardized testing (*Scoring High on the Iowa Test of Basic Skills*, 1987). They assert that their materials uphold the principles of the *Code of Fair Testing Practices in Education* (1988) by identifying learning gaps and removing sources of irrelevant difficulty by familiarizing children with test formats. This dispute can be framed in traditional terms of test validity but it also can be construed as a dispute about how learning occurs. Very likely, the antagonists differ in their beliefs about transfer of training from specific tasks, the role of practice and repetition, and the desirability of using multiple-choice formats for first-time instruction.

Lastly, the debate between Popham (1987) and Bracey (1987) or Popham and Shepard (1988) about the efficacy of measurement-driven instruction is motivated by conflicting learning theories. It is not just that we disagree about unintended side-effects of measurement-driven instruction, as when tested content grows to command more and more instructional time. Bracey, Shepard, and others disagree fundamentally with measurement-driven basic-skills instruction because it is based on a model of learning which holds that basic skills should be taught and mastered before going on to higher order problems, as Popham suggests when he

says, "Creative teachers can efficiently promote mastery of content-to-be-tested and then get on with other classroom pursuits" (p.682).

Although these examples document that differences of opinion exist about the role of testing, more thoroughgoing analysis is needed to examine whether these differences can be understood in terms of implicit assumptions about learning rather than some other value dispute such as differences in political goals for education. To undertake a more systematic examination of measurement specialists beliefs and their import for testing practice, this paper is organized into four parts:

1. An analysis of interview data from a nationally representative sample of 50 district testing directors.

2. A comparison of test directors' conceptions of learning with the frameworks of criterion-referenced testing, programmed instruction, and behaviorist psychology.

3. Consideration of a competing learning model from cognitive psychology.

4. Implications of explicit understandings of learning theory for reform of assessment practice.

Implicit learning theories: Interviews with 50 district test specialists

Data source. The interview transcripts examined here were collected as part of a larger study to replicate and extend Cannell's (1987) controversial report which asserted that all 50 states and 90 percent of U.S. school districts claim to be above average. Test data from the 35 states with normative statistics and from 153 districts (responding from a stratified random sample of 175) were reported in Linn, Graue, and Sanders (1990). The Linn et al. technical report also describes the method of sampling districts by region, size and socio-economic strata and includes the original survey instruments, both mailed questionnaires and telephone interview protocols. As described in Shepard (1990), telephone interviews were conducted with the directors of testing from all of the 50 states regarding the uses of test data, the process of test selection, time spent on teaching tested objectives, objectives given less time as a result of the test, guidelines for test preparation, typical and extreme practices in preparing students to take tests, and test security efforts and experiences. Parallel telephone interviews, which provided the data examined here, also were conducted with a subsample of 50 district test directors. Methods by which the district subsample was selected to be representative are described in Appendix E of Linn et al.

Data analysis. Although test directors' elaborations about the purpose of testing, and indirectly their assumptions about learning or instruction, sometimes occurred in answer to any of the interview questions, three prompts were selected for systematic reanalysis because these questions most often elicited talk about the effects of testing on instruction and learning. As shown in Table 1 (see Appendix A), questions 15, 16, and 17 asked whether efforts had been made to ensure that the curriculum and district (or state) test were aligned, whether teachers spend more time teaching the specific objectives on the tests than they would if the tests were not required, and whether important objectives are given less time or emphasis because they are not included on the test.

After responses to questions 15-17 were read separately and counted yes, no, or don't know, interview transcripts for the question sets were reread and characterized by a phrase or sentence to reflect each respondent's overall opinion about the effect of mandated tests on instruction in the district. Similar responses were then grouped together to form categories. To facilitate the initial sorting task

(i.e., to check for similarity within category and meaningful distinctions between categories) and later as a reporting device, categories were arranged along a continuum from least to greatest test influence on instruction. Although the initial reading and summarization of state interviews (Shepard, 1990) had suggested two other possible categorization schemes (views about criterion-referenced testing and learning or positive versus negative opinions about testing impact), the decision was made to organize the data in terms of the degree of instructional influence of tests—this scheme stayed closest to the survey questions as posed and therefore required the least inference on the part of the coder. This continuum also accounted for all of the data, whereas the other schemes left some cases which could not be accurately categorized. In keeping with the decision to stay close to the data for initial analysis, responses were located on the continuum according to the explicit answer choices of the respondents. Often a test director would describe a situation which implied substantial influence of tests on instruction to the interviewer or to the reader; nonetheless, efforts were made to categorize responses from the perspective of the respondent. This procedure sometimes led to different categorizations for highly similar accounts. For example, in Table 1, test director 9 in Category II and test director 15 in Category IV gave very similar answers about the tendency for teachers to pay attention to tested objectives and about district efforts to make sure that teachers attend to important objectives beyond those tested. They differed, however, in their explicit answers to question 16, with only one saying that more time was spent teaching tested objectives, and were therefore assigned to different categories.

Quantitative and qualitative data displays were developed. Brief phrases were used to convey different meanings for yes-no responses. Paraphrased quotations were developed to represent the gist of each category. Then shortened quotations were selected to provide specific examples of the types of answers given in each category.

Inferences about implicit learning theories. Clearly, measurement specialists in these two samples were not asked directly about their beliefs or theories of learning. Inference is required to hear assumptions about learning in talk about the effects of testing. Although this mode of investigation is not as concrete as some would like, it is customary to use indirect means to study the implicit theories of practitioners given that non-experts are not expected to have their theories easily accessible to report in propositional form. (Although test directors have expertise about measurement, they do not usually consider themselves to be experts about learning theory.)

Interpretations about what measurement specialists believe about learning are based on reanalysis of the primary narrative data. Again, descriptive codes were used to typify the responses. Those codes eventually became the propositional summaries used here to present the data. The data were reread for counterexamples. In general, the data did not produce equally elaborated competing theories of learning. Instead, the dominant model which seems to be widely shared in the profession is one which we called "the criterion-referenced-testing learning theory." A competing perspective, much less well elaborated in terms of an underlying learning model, might be called "the anti-measurement-driven instruction" position. As stated previously, some cases could not be categorized accurately at this higher level of inference. Therefore, beliefs about learning are presented below as propositions followed by supporting quotations and estimates of the proportion of cases accounted for. The first two propositions characterize the criterion-referenced-testing learning theory perspective. By way of contrast, the third proposition summarizes the more loosely defined anti-measurement-driven instruction position.

1. If a test is "criterion-referenced" or "curriculum-referenced," it is desirable for instructional effort to be redirected toward the test. The term criterion-referenced test is in quotation marks because test directors often referred to tests keyed to important instructional objectives as representing the appropriate goals of instruction even when they were off-the-shelf standardized norm-referenced tests. Thus I am using the term to characterize their way of speaking about the use of a test matched to important objectives even though sometimes they did not use the term explicitly. Two entire categories of responses on the instructional effects dimension in Table 1 can be thought of as "criterion-referenced-testing" types, Category III and Category V. Both groups reported a great deal of instructional effort addressed to tested objectives and emphasized that these were the important objectives that should be taught. Respondents in Category III, however, denied that this focusing required any redirecting of attention from what would have been taught if the test were not used.

Criterion-referenced-testing rhetoric is epitomized by respondent III.10 (Category III, response #10):

> We have a locally developed criterion referenced testing program, and these are skills that we have identified as being absolutely essential, and we test and retest until students show mastery. This is the kind of test that we think teachers should teach to, not particular items and answers of course, but really focus on the curriculum, because we have identified [] as key.

In other words, the tests and the curriculum are synonymous. Test director III.11 speaks in the same criterion-referenced terms about the standardized norm-referenced test in use in his district for the past 10 years:

> [16 (More time teaching tested objectives?)]
>
> No. I think that most of the skills that are appraised in the assessment instruments are part of our curriculum. They've always been part of the curriculum. When we're talking about skills, they've been there. I think pretty much the assessment instruments match what skills have been taught and are being taught.

Likewise, any of the quotations in Category V can be used as examples of a learning model, which says something like the following: "In order for children to learn effectively in schools, the schools must have a well-specified set of objectives, accountability tests should be keyed to these essential skills, and feedback should be provided about how well students have mastered the desired objectives." For example, respondent V.20:

> [16 (More time teaching tested objectives?)]
>
> Probably. They don't teach to items. We don't give them item analysis. We give them an integrated report grouped by domain. For example, for dealing with reading comprehension, we would have broken that down through a computer to facts or opinion, to main idea, to details, to sequence, to generalization. They would not see individual items. So they teach to those areas. Those areas, in turn, are curriculum referenced, and there are support materials for all of them.

Categories III and V account for 28 percent of all the district test directors. In addition, approximately half of the respondents in Categories IV and VI also gave positive accounts of a test carefully matched to the curriculum which improved instruction by directing attention to important objectives. Thus, half of the district

test directors in the national sample subscribe to the "criterion-referenced-testing" learning model.

Although answers to the question on test-curriculum alignment (#15) were included in the initial analysis, they were not often used as the basis for categorization and are only occasionally included in the sample quotations in Table 1. However, both the quantitative summary and a separate reading of question 15 data lend additional support to the conclusion that approximately half of all district test directors have a "criterion-referenced" view of testing and learning. In Table 1, 41 of the 50 district test directors answered "Yes," that efforts had been made to assure that the curriculum and test were aligned. Of these, the first 15 (30 percent of all test directors) answered primarily in terms of the test, usually a standardized norm-referenced measure, being selected to match the curriculum. Although the test selection process could later shape instruction if there were a great deal of emphasis on the test, these answers seemed to be framed in more traditional terms regarding the "content validity" of the test and were not considered necessarily as evidence of a criterion-referenced perspective. The remaining 26 test directors, however (52 percent of all the respondents), described much more extensive efforts to bring curriculum and teaching in line with the test, treating the test as the appropriate and desired focus for instruction. In addition to the criterion-referenced viewpoint of respondents in Categories III and V, the following quotations are answers to question 15, selected to represent those who espouse a criterion-reference view of test-curriculum alignment from among respondents in Categories IV and VI. (Original identification codes are used when the case was not one of the illustrative cases in Table 1.)

> IV.13. Those are our curriculum-referenced tests. There are curriculum guides for all of the major areas, reading, math and science and social studies, and we identified, in conjunction with the office of curriculum and instruction, key objectives that should ideally be mastered by the end of a given year; and that's how the content of the tests were specified....And of course the curriculum-referenced tests measure the curriculum and then we have done correlations between our curriculum tests, measuring our curriculum, and the Metropolitan test.
>
> IV.14. Yes, very extensive. With regard to the state test,...there was a major effort to do a curriculum match between the content of the state test and the curriculum of the school district.
>
> IV.[1841]. If I can use a term that's often used by [], we are very much involved in a test-driven curriculum, right or wrong. As we look at what the tests are attempting to measure, we have made adjustments in our curriculum to make sure that those pieces are in fact being covered.
>
> VI.32. Yes, there have been strands and objectives which have been prepared for [city] which would identify those strands and objectives which are measured by the CAT, also by our [state] test. So there would be correlations that have been developed for both of these tests to identify those areas and to provide techniques or lessons or methods that would help teachers obtain these objectives in classes.
>
> VI.33. There's been a lot of initiatives and reform legislation from the state, which has caused the instructional people to revise curriculum. When those have been revised, and this is what I'm told by the instructional people, they look at the CTBS test objectives, the state assessment test, the performance standards that the state has set in certain skill areas and subject areas, and then also the textbooks that we've adopted and try to get the curriculum in line with all of those areas. But they certainly pay attention to the testing.

VI.43. We developed a whole new technique of looking at the item analysis so that, instead of saying that on item 13 you did poorly, we would get into descriptive phrases and illustrate clusters of items that might be measures of the same skills....The curriculum people were able to look at a set of skills where you're consistently low across the years. The curriculum people were charged with the responsibility to look at whatever materials might be developed to help the schools to make sure that they were at least addressing the concepts and skills appropriately.

To restate then, test directors who think about learning from a criterion-referenced-testing perspective believe that it is appropriate and desirable for the test to be the target for instruction. This perspective is shared by half of the sample of district test directors, many of whom were describing a local or state use of a norm-referenced test rather than a test designed specifically as a criterion-referenced measure.

2. Basic skills are the most important learning goals, especially for elementary education, because basic skills are the building blocks or prerequisites for subsequent learning. Instances of the "basic skills" proposition were less frequent and tended to be embedded within the protocols already associated with proposition 1. The following excerpts are illustrative of the perspective that learning objectives should be sequenced to ensure mastery.

V.19. But if you're attempting to ready kids for the achievement test, you're attempting to ready students for the curriculum tests that are developed within the local efforts. Then that could take most of the time....But when you say less important (question 17), I don't know. The things that we try to stress are what is important. And of course you have terminal objectives and supporting objectives. But to push the terminal objectives which one might consider important, you have to in many respects touch upon the building-block objectives.

V.27. Well, it is a criterion-referenced test, the [State test] that I mentioned, and all of those skills are remediated, taught and then remediated after the test at every grade level, and that is its purpose, because by the time they get to be in high school prior to graduation, they must have mastered them. In order that the courts would allow us to withhold a diploma, we had to give evidence that we are teaching those skills adequately.

V.28. We have what we call the basic elements of our curriculum, and our [Local tests] reflect those basic elements. (State test aligned?) As closely as we can get it. That sometimes is a problem, but by and large, the state has made quite an effort in the last four or five years to get everybody in line for at least minimum skills or basic skills....I don't believe the test eliminates any really important objectives.

Occasionally, respondents who had not previously been classified as having a criterion-referenced testing perspective referred to the importance of teaching essential skills. For example:

IV.16. So they established this list of essential skills. It took about a year to do that for each grade and each of those subject areas, what ought to be taught, the essential skills that ought to be taught at each grade level. And once we received these, we made sure that every teacher and administrator in our district had a copy of these, and they were instructed to make sure that they taught all of these essential skills at their particular grade level.

Together, propositions 1 and 2 comprise what I have called the criterion-referenced-testing learning theory. These themes or shared understandings, which seemed to recur in the first reading of the data, were the impetus for this paper. More systematic investigation confirms that many measurement specialists have a coherent view of learning as the sequential mastery of basic skills. Testing is closely tied to instruction because it assesses what students know and don't know in their progress toward mastery. This underlying learning model is elaborated further in the next section of the paper by examining the work of psychologists from whom measurement specialists appear to have drawn their assumptions about learning.

To complete this second-level analysis, however, where learning theories are inferred from narratives about instructional effects of testing, I offer one final belief or proposition which accounts for most of the cases not characterized by the criterion-referenced-testing perspective.

3. Tests should be for monitoring but should not drive instruction. As stated previously, whatever learning beliefs are held by those who do not believe in the criterion-referenced-testing learning theory, they were not adequately elicited by these indirect questions on the instructional effects of testing. That is, in the course of telling whether they believed that tests in their jurisdiction had or had not increased the amount of time spent teaching specific objectives, they did not reveal as much about their learning theory as the criterion-referenced group had. Perhaps this asymmetry in the adequacy of the data occurred because large-scale testing and learning are closely tied together only from the perspective of the criterion-referenced-testing group. Thus, whether direct or indirect, a different line of questioning would have been necessary to elicit responses that would reveal the implicit learning theories of specialists not in the criterion-referenced-testing camp.

Other view points held by this last group of testing directors, at least about the role of testing in instruction, are represented reasonably well by returning to the first level of analysis summarized in Table 1. Respondents in Category I describe testing situations where very little instructional attention is given to tested content *per se*: "What's on the Iowa Test really does not determine what's going to be taught in the classroom" (I.3). And generally they appeared to think it was a good thing that tests do no have an undue influence on teaching. By implication, test directors in Category II also do not approve of having the test be the exclusive target for instruction, because they each described mechanisms that ensure that the entire curriculum is taught, not just what's tested. Similarly, some members of Category IV and Category VI appear to reject the idea of targeting instruction by means of the test. For example, according to test director IV.13, "I think the issue is with teachers who are not as seasoned. For them in particular, tests circumscribe the curriculum and determine it." Several of the respondents in Category VI, those who did not espouse a criterion-referenced perspective, conveyed a negative tone. This last group of district test directors seems to believe that some important objectives are given short shrift because they are not tested. As noted by director VI.41, "We do have some evidence that shows when you have a basic skills test as we do statewide that the amount of effort that goes into that does subtract from some of the higher level skills." However, none of the test directors who gave slightly negative responses about the effects of testing on instruction mentioned being concerned about basic skills testing *per se* or complained about the sequencing of instruction to ensure mastery of basics skills first. Rather, they seemed to be concerned that emphasis on testing had given basic skills disproportionate weight compared to unmeasured skills.

From this point in the paper onward I focus only on the dominant model of learning held by measurement specialists, setting aside the viewpoints of those in this last group who seem to be against measurement driven instruction. The next section of the paper is intended to illustrate the origins of the criterion-referenced-

testing perspective in behaviorist psychology. Although the third section of the paper introduces a cognitive or constructivist perspective in contrast to behaviorism, there is no implication intended that these learning theories underlie the thinking of a significant group of measurement specialists. It seems more likely to me, from a sense of the data too vague to document, that this "other" group of measurement specialists holds to older views of measurement, relying on concepts of construct validity and sampling from a domain of content, but without a professionally shared theory of learning. (Note that traditional psychometrics comes from the psychology of individual differences which does not address the mechanisms of learning.)

Origins of Measurement Specialists' Learning Theory in Programmed Instruction and Behavioral Psychology

How is it that so many measurement specialists talk in such similar terms about the sequencing of student learning and the close alignment of tests to instruction? Several explanations are possible. It is conceivable that there is only one true way to organize effective instruction, and measurement specialists all arrived independently at the same conclusion. It is more likely, however, that measurement specialists who share very similar views about learning had the same training in educational psychology or adopted these views implicitly when they adopted the principles of criterion-referenced testing. Most likely some combination of these explanations is at work.

My purpose here is to argue that the criterion-referenced-testing paradigm is grounded in the learning theory of behaviorism (and before that in Thorndike's connectionism), and that implicitly the majority of measurement specialists invoke this model when they think about learning. My treatment of behaviorism is necessarily simplistic, focusing on the principles that parallel those in the accounts of measurement specialists and ignoring other major aspects of the theory such as the contingencies of reinforcement. I also gloss over disagreements among behaviorists about theoretical details and their implications for instruction. I am trying to describe what contemporary measurement specialists remember from behaviorism, not the fully elaborated positions of the original thinkers.

Table 2 is an historical data display of quotations intended to exemplify the learning and instructional model of behavioral psychology. Whether couched in terms of teaching machines, learning hierarchies, programmed learning, mastery learning, or criterion-referenced testing, these authors share the same learning theory. This theory can be organized into two principles which correspond to the criterion-referenced-testing propositions in section 1. I will summarize these principles, but in reverse order. Not surprisingly, the learning proposition comes first in the discourse of the psychologists and the testing-instruction principle comes second.

1. Learning is seen to be linear and sequential. Complex understandings can only occur by the accretion of elemental, prerequisite learnings. In Skinner's (1954) words, "The whole process of becoming competent in any field must be divided into a very large number of very small steps, and reinforcement must be contingent upon the accomplishment of each step" (p.94). And according to Gagne (1970), "Thus, it becomes possible to 'work backward' from any given objective of learning to determine what the prerequisite learnings must be—if necessary, all the way back to chains and simple discriminations" (p.242). The whole idea was to break desired learnings into constituent elements and teach these one by one.

This view of learning is captured visually by pictures of learning hierarchies. For example, in Figure 1 (see Appendix B) we see two hypothetical sequences offered by Glaser and Nitko (1971), one simply linear and one where several

streams of prerequisites are essential to higher, terminal objectives. Real attempts to define the hierarchies of objectives essential to the acquisition of particular skills and concepts are represented by the following examples from Ferguson (1969) and Gagne (1970), Figures 2-4. The implications of this model for instruction are conveyed best by Madeline Hunter's metaphor of a brick wall (i.e., it is not possible to lay the bricks in the fifth layer until the first, second, third, and fourth layers are complete).

Given the specificity and minuteness of these analyses, one can imagine a highly complex set of instructional maps needed to address all the subject matter goals of public education (see hypothetical example in Figure 5). Although many prerequisite strands may be acquired in parallel, nonetheless the hierarchical and sequential nature of learning within strands is insisted upon. As an aside, I might note that the image of parallel learning strands, each sequentially ordered and marked by essential milestones, is also consistent with the public's understanding of the immutability of grade level achievement, requiring grade retention as the only remedy to deficient skill acquisition (Shepard & Smith, 1989).

Perhaps the most serious consequence of the programmed learning or mastery learning model of instruction is that higher order skills, which occur later in the hierarchies, are not introduced until after prerequisite skills have been mastered. When Resnick and Resnick (in press) explained the inadequacies of associationist and behaviorist theories, they described the assumptions of decomposability and decontextualization. The model assumes that component skills can be adequately defined and mastered independently and out of context. Only then are more advanced thinking skills acquired by "adding up" or assembling component abilities.

2. To facilitate learning, assessment should be closely allied with instruction. Tests should exactly specify desired behavioral outcomes of instruction and should be used at each learning juncture (i.e., one should "test—teach—test"). Principle number 2 in the behaviorist learning model corresponds to proposition 1 in the criterion-referenced-testing implicit learning model held by measurement specialists. The important role of testing to judge progress in mastery learning is exemplified by several quotations in Table 2.

> In practice, implementation of a mastery curriculum implies that children will be permitted to proceed through the curriculum at varied rates and in various styles, skipping formal instruction altogether in skills or concepts they are able to master in other ways. This demand for individualization, in turn, requires that there be some method of assessing mastery of the various objectives in the curriculum (Resnick, Wang, & Kaplan, 1973, p. 700).
>
> Given our description of the learning tasks for <u>each</u> unit, we have then constructed brief diagnostic-progress tests to determine which of the unit's tasks the student has or has not mastered and what he must do to complete his unit learning (Bloom, 1971, p. 58).
>
> When a student has completed a prescription, he is tested. The test is corrected immediately, and if he gets a grade of 85 percent or better he moves on to a new prescription assigned by the teacher. If he falls below 85 percent, the teacher offers a series of alternative activities to correct weakness, including special individual tutoring. He is not permitted to advance to a new unit of work until he achieves the 85 percent proficiency rating (Education U.S.A., 1968, p. 4).

Taking principles 1 and 2 together, it should be clear that the behaviorist and programmed learning model also relies on assumptions about the nature of tests.

First, it assumes that all important learning objectives can be specified and measured both completely and exhaustively. Each of the learning steps is small enough that highly homogeneous tests can be used to measure mastery at each step without inference to some broader set of test questions. The items for a particular objective are not thought to be sampled from a larger domain, nor is it expected that any aspect of the objective is left unassessed by the item set. If students can do what the questions ask, they have fully mastered the objective. Because each set of test items is a perfect instantiation of the learning objective, highly similar items can be used to test and retest without harm to the integrity of the measurement. It is also assumed that all learning steps will be measured exhaustively at least for instructional purposes. The only circumstances where the behaviorist model admits of the need for item sampling—and therefore inference or generalizability beyond the actual test questions administered—is for review tests or placement tests, where a sampling of some of the items from some of the objectives is permitted. Even here, however, the exhaustive specification of objectives and their explicit sequencing make the process of inference a mechanical one. It is not considered possible in this low inference system to function well on the test and not have fully mastered the intended skills and concepts. Just as measurement specialists in the first section gave answers that treated the test and curriculum as synonymous, it should be clear from the behaviorist perspective that tests and learning objectives are equivalent and, therefore, that teaching to tested objectives is synonymous with good instruction.

A Competing Learning Model from Cognitive and Constructivist Psychology

But what if learning is not linear and is not acquired by assembling bits of simpler learnings? What if the process of learning is more like a Faulknerian novel where one has glimpses and a vague outline of ideas before each of the concrete elements of a story are fit into place? What if learning is more like an image gradually brought into sharper focus as the learner makes connections, not stimulus-response connections but connections and relations among ideas? Or what if learning is like a mosaic with specific bits of knowledge situated within some larger design? But even these metaphors are wrong; they imply that a knowledge structure external to the student is exactly what is reproduced and cemented inside the student's head, whereas we know that learning requires reorganizing and restructuring as one learns. A more organic conception is needed.

In contrast to the linear pictures presented earlier, consider the following examples. Figure 6 is a semantic network drawn to display one child's concepts and connections after a lesson on two-digit subtraction with regrouping (Leinhardt, 1989). Figure 7 is also a semantic network representation to show the organized knowledge a 4 1/2 year old boy had of dinosaurs and their classification (Chi & Koeske, 1983).

Contemporary cognitive psychology has built on the very old idea that things are easier to learn if they make sense. We can think of learning as a process whereby students take in information, interpret it, connect it to what they already know, and if necessary reorganize their mental structures to accommodate new understandings. Learners construct and then reconstruct mental models that organize ideas and their interrelation. Because I am a novice in trying to understand cognitive psychology, let me quote a richer description by Glaser (1984).

> When schema knowledge is viewed as a set of theories, it becomes a prime target for instruction. We can view a schema as a pedagogical mental structure, one that enables learning by facilitating memory retrieval and the learner's capacity to make inferences on the basis of current knowledge. When dealing with individuals who lack adequate knowledge organization, we must provide a beginning knowledge structure. This might be

> accomplished either by providing overt organizational schemes or by teaching temporary models as scaffolds for new information. These temporary models, or pedagogical theories as I have called them, are regularly devised by ingenious teachers. Such structure, when they are interrogated, instantiated, or falsified, help organize new knowledge and offer a basis for problem solving that leads to the formation of more complete and expert schemata. The process of knowledge acquisition can be seen as the successive development of structures which are tested and modified or replaced in ways that facilitate learning and thinking (p. 101).

As an example, think about learning the measurement concepts of reliability and validity. If we had a strictly linear idea about how these ideas are acquired, we might focus on mastery of prerequisite knowledge such as the standard deviation, normal curve, and the correlation coefficient. From the perspective of cognitive psychology, however, students come to the learning of these measurement concepts with a great deal of prior knowledge having to do with their own experiences taking fair and unfair tests. Students begin with undifferentiated equivalences between good, fair, reliable, and valid tests, and ones they do well on. Good instruction is aimed at eliciting prior understandings and explicating the congruence or misfit between technical definitions and everyday conceptions. As noted by Glaser (1984), the progression is from simpler mental models to more complex ones, rather than a progression from facts to comprehension to analysis. The first pass at textbook learning creates a mental image where reliability and validity are two equally important side-by-side constructs, as illustrated in Figure 8. Then as understanding develops, the major concepts are transformed, subordinate and superordinate concepts are recognized, hierarchies emerge, and bits of information are located in the meaning network. Figure 9 represents a more elaborated, expert view, revealing my own understandings of the interconnections among reliability and validity and other measurement concepts. The evolution and restructuring in my conceptual network is obviously influenced by the expanding definition of validity in the professional literature over the last two decades (see the Test Standards [APA, 1985] and Messick [1989]).

This major principle of cognitive psychology, that learning occurs by the individual's active construction of mental schemas, applies even to the youngest children. All learning requires us to make sense of what we are trying to learn. To quote Lauren and Dan Resnick (in press):

> One of the most important findings of recent research on thinking is that the kinds of mental processes associated with thinking are not restricted to an advanced or "higher order" stage of mental development. Instead, thinking and reasoning are intimately involved in successfully learning even elementary levels of reading, mathematics, and other school subjects. Cognitive research on children's learning of basic skills reveals that reading, writing, and arithmetic—the three Rs—involve important components of inference, judgment, and active mental construction. The traditional view that the basics can be taught as routine skills, with thinking and reasoning to follow later, can no longer guide our educational practice (MS p. 4).

The Resnicks substantiate this claim with cognitive research from beginning reading and mathematics learning. In reading for example, comprehension of even simple texts requires inference on the part of the reader. Authors cannot stipulate every detail needed for understanding. Competent readers supply implicit meanings and interpret the text to themselves (tell themselves the story) so automatically that they are unaware of this process until they fail to comprehend. Then good readers have strategies to reread and interrogate the text until they do comprehend. Poor readers do not engage in this kind of active translation of text necessary to make

sense of it. Therefore, they often fail to comprehend even when they can satisfactorily decode every word.

Current research on learning has many more things to teach us about how students learn, and therefore about the organization of instruction and the nature of tests that would facilitate learning. In contrasting cognitive theory with behaviorism I have focused primarily on findings regarding cognitive structures and the notion that thinking comes before, not after, the acquisition of facts. Other fundamentally important findings have to do with the social aspects of learning (Resnick, 1987) and the move away from generic thinking skills to those embedded in particular knowledge domains (Glaser, 1984). To develop assessments more compatible with the cognitive view of learning would require overturning of what the Resnicks called the decomposability and decontextualization assumptions of older learning theories. Tests ought not ask for demonstration of small, discrete skills practiced in isolation. They should be more ambitious instruments aimed at detecting what mental representations students hold of important ideas and what facility students have in bringing these understandings to bear in solving new problems.

Conclusion: Implications for Measurement Practice

Three main points are made in the respective sections of this paper:

1. Based on qualitative analysis of interview data from a representative sample of 50 district testing directors, it is asserted that a majority of measurement specialists operate from implicit learning theories that encourage close alignment of tests with curriculum and judicious teaching of tested content.

2. These beliefs, associated with criterion-referenced testing, derive from behaviorist learning theory which requires sequential mastery of constituent skills and behaviorally explicit testing of each learning step.

3. The sequential, facts-before-thinking model of learning is contradicted by a substantial body of evidence from cognitive psychology.

My argument is that hidden assumptions about learning should be examined precisely because they are covert. What we believe about learning and the intended effect of testing on learning should be considered directly, not "smuggled in" by the adoption of a seemingly technically superior testing theory. What measurement specialists believe about learning does shape practice, including instructional practice. Although we have formal theories about test validity and formal means to evaluate how technical decisions affect the meaning of test scores, we do not have explicit ways to examine and debate our understandings of learning theory. Left unexamined, it is possible for 30-year-old theory to still have a pervasive influence. Note that in selecting quotations to characterize the behaviorist position in Table 2 I purposely chose examples from Glaser's Individually-Prescribed Instruction and Resnick's earlier work. Their work in the 1980's is nearly a repudiation, certainly a significant transformation of their earlier understandings. They have changed but we have not, primarily because it has not been our purpose to learn about learning.

Thus, I propose that we engage in formal debate about our theories and expectations for the effects of tests as well as considering the empirical evidence of these effects. There has been a tremendous hue and cry in this decade about the negative effects of high-stakes testing inaugurated by educational reform. Often the connotation is that the undesirable consequences of testing are unintended side-effects caused by poor implementation or perversion of desirable policies. It is possible, however, with greater theoretical insight, that we would see many of these effects as predictable, the direct consequence of what new theories of learning

would expect from old instructional practices enforced by the tests. Historically, psychometricians were psychologists and were, therefore, unlikely to lose touch with fundamental transformations in learning theories. As we attempt to develop alternative assessments we should be guided by a deep understanding of the teaching and learning context, not just our statistical models or the surface features of new tests.

References

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: Author.

American School Publishers. (1989). Code of fair testing practices in education. *TestNet, 1,* 1, 4.

Bloom, B.S. (Ed.). (1956). *Taxonomy of educational objectives: Handbook I, cognitive domain.* New York: David McKay.

Bloom, B.S. (1971). Mastery learning. In J.H. Block (Ed.), *Mastery learning: Theory and practice.* New York: Holt, Rinehart, & Winston.

Bracey, G.W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. *Phi Delta Kappan, 68,* 683-686.

Cannell, J.J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average* (2nd ed.). Daniels, WV: Friends for Education.

Chi, M.T.H., & Koeske, R.D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19,* 29-39.

Clark, C.M. (1988). Asking the right questions about teacher preparation: Contributions of research on teacher thinking. *Educational Researcher, 17*(2), 5-12.

Code of Fair Testing Practices in Education. (1988). Washington, DC: Joint Committee on Testing Practices.

Education U.S.A. (1968). *Individually prescribed instruction.* Washington, DC: Author.

Ferguson, R.L. (1969). *The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction.* Unpublished doctoral dissertation, University of Pittsburgh.

Gagne, R.M. (1965). *The conditions of learning* (2nd ed.). New York: Holt, Rinehart, & Winston.

Gagne, R.M. (1970). *The conditions of learning* (2nd ed.). New York: Holt, Rinehart, & Winston.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93-104.

Glaser, R., & Nitko, A.J. (1971). Measurement in learning and instruction. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Leinhardt, G. (1989). Development of an expert explanation: An analysis of a sequence of subtraction lessons. In L.B. Resnick (Ed.), *Knowing, learning, and instruction.* Hillsdale, NJ: Lawrence Erlbaum.

Linn, R.L., Graue, M.E., & Sanders, N.M. (1990). *Comparing state and district test results to national norms: Interpretations of scoring "above the national average"* (CSE Tech. Rep. No. 308). Los Angeles: UCLA Center for the Study of Evaluation.

Mehrens, W.A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice, 8,* 14-22.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: American Council on Education, MacMillan.

Popham, W.J. (1978). *Criterion-referenced measurement.* Englewood Cliffs, NJ: Prentice-Hall.

Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68,* 679-682.

Resnick, L.B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Resnick, L.B., & Resnick, D.P. (in press). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction.* Boston: Kluwer Academic Publishers.

Resnick, L.B., Wang, M.C., & Kaplan, J. (1973). Task analysis in curriculum design: A hierarchically sequenced introductory mathematics curriculum. *Journal of Applied Behavior Analysis, 6,* 679-710.

Scoring High on the Iowa Test of Basic Skills, Teacher's Edition, Book B. (1987). New York: Random House.

Scriven, M. (1967). The methodology of evaluation. In R. Stake (Ed.), *Perspectives of curriculum evaluation.* Chicago: Rand McNally.

Shepard, L.A. (1988, April). *Should instruction be measurement driven: A debate.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Shepard, L.A. (1990). *"Inflated test score gains": Is it old norms or teaching the test?* (CSE Tech. Rep. No. 307). Los Angeles: UCLA Center for the Study of Evaluation.

Shepard, L.A., & Smith, M.L. (Eds.). (1989). *Flunking grades: Research and policies on retention.* London: Falmer Press.

Silberman, H.F. (1965). Reading and related verbal learning. In R. Glaser (Ed.), *Teaching machines and programmed learning, II: Data and directions.* Washington, DC: National Education Association.

Skinner, B.F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24,* 86-97.

Skinner, B.F. (1965). Reflections on a decade of teaching machines. In R. Glaser (Ed.), *Teaching machines and programmed learning, II: Data and directions.* Washington, DC: National Education Association.

Appendix A

Table 1

Interview Responses of District Test Coordinators Regarding Test-Curriculum Alignment and Instructional Influence of Tests (n = 50)

QUANTITATIVE SUMMARY BY QUESTION:

---

15. Have there been district efforts to assure that the curriculum and the district test are aligned? [aligned with the state test?]

No	6	just studying that now. what's on the Iowa does not determine what we teach. content validity to select test but don't let it drive curriculum.
Yes	41	test selected to match curriculum. (12) but focus on our curriculum more. (2) but not making wholesale changes. (1) local curriculum must reflect state test. (15) test selected to match, then further alignment. (5) CRT test tailored to objectives. (3) customized test. (2) test driven. (1)
DK	3	

---

16. Do you think that teachers spend more time teaching the specific objectives on the test(s) than they would if the tests were not required? How much more time?

No	12	We follow our curriculum (rather than test). (5) The test matches our curriculum. (2) CRT, supposed to teach to objectives. (1) don't pay much attention to tests. (2) We monitor our teachers. (1) because test samples objectives each year. (1)
Yes	35	definitely. always more emphasis on what's tested. We encourage them to. because of how we give information back to them. as they get down to the wire, probably a lot more time. more than I would like. (See categorical summaries for more examples.)
Varies	1	
DK	1	
NR	1	

---

17. To what extent do you think important objectives are given less time or emphasis because they are not included on the test?

None	21	the test reflects our curriculum. the test is embedded in our curriculum. except for insecure teachers. teachers don't worry about the test. we monitor curriculum objectives. teach curriculum rather than test. teachers don't know the test yet.
Some	21	there has to be a trade-off. Yes, but these are the building-block skills. focus on the most important objectives. has more effect on sequence, to be sure it's covered before the test. especially for unseasoned teachers.
Varies	1	
DK	6	
NR	1	

EXAMPLES OF RESPONSES BY CATEGORY:

Note: After responses to questions 15-17 were read and counted yes, no, or don't know, the question sets were reread and categorized to reflect each respondent's overall opinion about the effect of mandated tests on instruction in the district. Each category is characterized by a paraphrased summary in boldface type. The number of responses in each category follows in parentheses. Categories are arranged here from least to greatest effects on instruction, according to the respondents.

Yes, No, and Don't Know responses to questions 15-17 are shown by letter abbreviations at the beginning of each quotation (e.g., YNDK). Question prompts [15], [16], and [17] are shown in text to indicate which question the respondent is answering in the selected quotations. Identification codes, reflecting region, size, SES, and replicate follow each quotation.

I. Teachers don't worry about tests. Focus is on curriculum. (7)

1. DKNN. ...[17] "I don't think there is any. Just because they don't appear on the test does not mean that they are not important, so we go ahead and teach them....People don't generally have access to those tests to know that the metric system isn't on the test, so why teach it?" [2131]

2. YNN. ...[16] "No because we have our curriculum. That's the forefront. We look at the curriculum and establish our requirements based on what we feel should be taught to children. When we make our curriculum we're looking at the state course of study. So our curriculum is closely modeled after the state course of study. [17] I think that's secondary. Maybe in some systems it becomes a primary objective, but

in our system it has stayed secondary, because we feel we have a good core curriculum. We feel pleased with what the state has established as its course of study and then our curriculum reflects that. And if it happens that that's also on the test, well and good" [3722].

3. YNN. ...[17] "To be honest with you, I don't think that our district or individual teachers look at the test that closely so that would not be a factor in their teaching. I would say that what's on the Iowa Test really does not determine what's going to be taught in the classroom" [4111].

4. YNN. ...[16] "Quite frankly, the teachers in our district don't pay a whole lot of attention to teaching to the test. They think that the test just serves a certain purpose and it only measures about 40 percent of what they teach anyway, so they don't worry about it. They just go ahead and teach and aren't really that worried about it" [4331].

5. YNN. ...[16] "I don't think so. I doubt that they're letting the tests drive them that much because in some of our analyses, we find that items tested may not be taught until later and some of our staff members have come up and said, 'I do not feel like our kids are ready for that until this point in time, so I am not even going to introduce that. I can introduce it at the time they are going to get it. I am not going to teach it just because it is going to be tested.'..." [4451]

II. Efforts to ensure focus on curriculum, not test. (5)

6. YNN. ...[16]..."They understand that it only covers a sample of the objectives in the curriculum...and they know that the objectives covered will change from year to year and so there is not a particular way they could move other than to say we now have a testing program that really measures our curriculum, therefore, we better be sure we teach our curriculum....[17]. I think there is definitely an emphasis. I mean even in test preparation, people go over test format with the kids and the schools certainly gear up for the test. You know, they know the test is coming and we do workshops on how to sort of incorporate test taking skills and your regular instruction, not just to give item after item for kids to practice on, but have kids make up questions during the course of the year...." [1822]

7. YNN. ...[17]..."I'm not sure. I would guess that probably not too much. I suppose there could be some instances where that would occur, but in general, we have a curriculum for our schools set up and they're expected to pretty much follow that curriculum. Our curriculum specialists and supervisors are out in the schools, and I would expect that that wouldn't be a real problem" [2731].

8. YNN. ...[16] "They might have some. But for the most part I would say no. You're going to have some who are going to want to look good, who might feel insecure. New teachers, things of that nature, might want to make sure that they cover the objectives that will be tested. But for the most part I don't think they're doing that at the expense of other more important things that need to be taught. And that's one of the things that we stress at our inservice activities, that the test items or test objectives (should not) dictate what you teach students" [3742].

9. YNN. ...[16]... "And I'm sure that there are individual teachers out there who might do that a few weeks before the test....But I don't think that that is a wide spread practice in the district for a couple of reasons: 1) We have an extensive teacher assessment program in the district, and it's a state required assessment

program....There is extensive observation of the teachers in the classroom. We have the essential elements that are required. Every content area has its lists of proficiencies and essential elements that are to be covered that year. There is a high level of accountability in a sense of what teachers are supposed to be doing in the classroom. Now, that's probably only going to be as good as the principals in the school, as so on, but I don't believe that this notion of teaching to the test and spending more time on these objectives is the wide spread practice in the schools" [4831].

III. Important objectives aren't slighted because test and curriculum are well matched. (3)

10. YNN. "We have a locally developed criterion referenced testing program, and these are skills that we have identified as being absolutely essential, and we test and retest until students show mastery. This is the kind of test that we think teachers should teach to, not particular items and answers of course, but really focus on the curriculum, because we have identified [ ] as key. In some respects, the district has put an inordinate amount of attention on achievement test results, and I can see why teachers or staff are inclined to focus on them" [1241].

11. YNN. ...[16] "No. I think that most of the skills that are appraised in the assessment instruments are part of our curriculum. They've always been part of the curriculum. When we're talking about skills, they've been there. I think pretty much the assessment instruments match what skills have been taught and are being taught" [1831].

12. DKYN. ...[16] "We don't give them any objectives on the tests. For the SAT, we don't publish any objectives from it. The SAT is a blind administration. For the State tests, they're supposed to teach the objectives because it is a criterion referenced test, and the State Department of Education distributes the objectives to each and every teacher. [17] All objectives are taught" [3835].

IV. Yes there is an emphasis on tested objectives, but these objectives are embedded in the curriculum. (9)

13. YYY. ...[17] "Yes, I do feel that there are some areas that are eliminated, not by a seasoned teacher so much, because I think a seasoned teacher who has a well run classroom and is knowledgeable about the curriculum will teach irrespective of the test, although is aware of the test, and is aware of the objectives, but still teaches what children need to know, and teaches what needs to be measured. I think the issue is with teachers who are not as seasoned. For them in particular, tests circumscribe the curriculum and determine it" [1722].

14. YYY. ...[16] "Yes. [17] To some extent. I would say that this school district has over the years attempted to integrate state mandated and county mandated testing into the instructional program, but that testing does not drive the curriculum." [1741]

15. YYN. ...[16]..."I think like in any other system, once you institute a testing program, there are people who are going to look at the objectives of the test and incorporate that into their instructional program....[17] In our elementary schools, we have an instructional management system to try to ensure that teachers cover important objectives" [3831].

16. YYN. ...[15]..."So they established this list of essential skills. It took about a year to do that for each grade and each of those subject areas, what ought to be taught, the essential skills that ought to be taught at each grade level. And once we received these, we made sure that every teacher and administrator in our district had a copy of these, and they were instructed to make sure that they taught all of these essential skills at their particular grade level. [16] I think in our district, they probably spend a little bit more time on this, but we never did make an official correlation between our curriculum and the [State] essential skills. We never did that, purposefully in a way. Because we didn't consider it worth our time, number one, and number two, we did not want to get into a situation where we put so much emphasis on this that teachers were actually being imprisoned by the state mandated testing program, and either teaching the test or teaching things that were really close to what was on the test" [3241].

17. YYN. [15] "Yes. The objectives have been correlated to the curriculum. (State test?) The standardized test is the state selected test. [16] Yes. (How much more time?) I couldn't tell you that. Well, first of all, the objectives of the test are for the most part imbedded in the curriculum, so they would be teaching the curriculum. But I think the emphasis is on...[what's tested.] When they get to the part of the curriculum or a skill in the curriculum that is going to be tested, then they give it more emphasis certainly, because what's tested is what's given emphasis" [3711].

18. YYN. ...[15] "No, not an effort to change the curriculum. We made an effort of correlating the two so we know where the gap is....We still have our own curriculum, but I think people have felt like they need to know what is on the test, the objectives. Now we make sure that everybody knows the objectives, that is published by Stanford, but I don't think the curriculum people have made any effort to really revamp the curriculum. [16] Yeah, I am sure they do. I am sure, if they know that an objective is on the test, and may even know the items on the test, obviously, the items are the same items and have been for four years...when they know that is on the test, they are going to make sure that it is covered" [3731].

V. Yes test focuses instruction, but these are the important objectives. (11)

19. YYY. [16] "I think they do give added emphasis to what's on the test. In a way, we foster that feeling by making available to the teachers, I call it a 'bullet sheet,' but it is a listing that CTB offers and lists all of the 90 objectives for the test. We do push one of their reports called 'The Category Objectives Report.' It shows how well students performed on various objectives. It lays out content a little more specifically than when you just say our total reading scores, main idea, literal recall, and so forth. We push that information and use of the information. [17] You can only put so much in the 'x' amount of time the teachers have. And there are a number of tests that we administer. We give our own curriculum tests. A lot of the curriculum based tests do have overlap on the standardized achievement test. But if you're attempting to ready kids for the achievement test, you're attempting to ready students for the curriculum tests that are developed within the local efforts. Then that could take most of the time....But when you say less important, I don't know. The things that we try to stress are what is important. And of course you have terminal objectives and supporting objectives. But to push the terminal objectives which one might consider important, you have to in many respects touch upon the building block objectives" [1731].

20. YYY. [16] "Probably. They don't teach to items. We don't give them item analysis. We give them an integrated report grouped by domain. For example, for

dealing with reading comprehension, we would have broken that down through a computer to facts or opinion, to main idea, to details, to sequence, to generalization. They would not see individual items. So they teach to those areas. Those areas, in turn, are curriculum referenced, and there are support materials for all of them. [17] If it's not included on the test, then we have no handle on the extent to which people pay attention to it. In the elementary [grades], the focus is basic skills, so that the focus is very much on the kinds of measures that are there which are directly related to being able to read or directly related to being able to do computations and problem solving in mathematics. I mean, it's the same as the curricula" [1811].

21. YYY. [15] "Our test is primarily criterion referenced....We provided [the contractor] with a series of objectives and they provided us with anywhere from four to eight items with national standardization information on each of the items....[17] I would say that the process helps us to guarantee that the most important objectives are being taught and tested. But it's the nature of the beast. That means that there are certain other things that are not being taught, and there is nothing you can do about that" [1821].

22. YYDK. [16] ..."We do know that they are spending more time teaching those objectives, but again to clarify that, it's my feeling based on our staff development program and the sessions with those teachers involved that they are devoting more time to objectives that are measured by the tests where student performance needs to be improved" [2551].

23. YYDK. [16]..."Yes they do, and that's particularly true because of the criterion referenced test. For most of us, that's an intended outcome. I'm not sure it's so much more time spent on particular things as it is [that] they now organize what they present to kids in a slightly different way. They sequence instruction a little differently now because they're matching the way the course has been structured and the order in which we're going to be testing those kinds of things" [2732].

24. YYDK. [16] "I would say yes. As I said, we have competency testing and this is based on the local objectives of the curriculum, and those teachers really do a very detailed job of teaching the objectives....[17] The way our objectives are arranged is that it seems like every objective is given the same weight in importance....Now we all know that there are some objectives that are more important than the others. But the teachers treat those darn objectives as if they were all equally important, and that is one of our problems. Even a minor objective is given the same weight as say finding the main idea" [2722].

25. YYY. [16] "They would probably teach the objectives anyway, if it's part of the local curriculum. That's an interesting question. The objectives tie into the state objectives which are supposedly measured on the state achievement tests. I know the prevailing attitude among the people in curriculum is that if the kids aren't tested on something, those teachers out there aren't going to teach it, and I don't know the extent to which that's true" [2831].

26. YYDK. [16] "Yes. As a matter of fact we encourage them to. When areas of the test do not have particular content validity for our curriculum, then we say, 'look this is on the test and you are not covering it in your class. Would you consider teaching this at this level?'" [3351]

27. YYY. [15] "Well, it is a criterion referenced test, the [State test] that I mentioned, and all of those skills are remediated, taught and then remediated after the test at every grade level, and that is its purpose, because by the time they get to be in high school prior to graduation, they must have mastered them. In order that the courts would allow us to withhold a diploma, we had to give evidence that we are teaching those skills adequately....[16] I don't think there is any doubt....but on the other hand, I'd like to think that it is a genuine effort to improve curriculum....[17] One of the mandates in the new test committee is to find a test that does have some higher order thinking skills on it. That is one of the things that the district is examining, and of course, that is one of the newest developments as I see it, in all the tests now they are talking about higher order thinking skills to be incorporated in achievement tests, to give people at the top to stretch a little bit more" [3531].

28. YYN. [15] "Oh yes! That's top priority. We have what we call the basic elements of our curriculum, and our [Local tests] reflect those basic elements. (State test?) As closely as we can get it. That sometimes is a problem, but by and large, the state has made quite an effort in the last 4 or 5 years to get everybody in line for at least minimum skills or basic skills. [16]...Of course, they don't know the test items, so that they can't teach to any of the test, but they are very aware of the kinds of things that are going to be done, and so they do stress it. I'm sure. [17] I don't believe the test eliminates any really important objectives" [4832].

29. YYN. [16] "Oh, I think it has considerable influence. I think that in the past there may have been some objectives that were never taught, and so now with an accountability factor [they are taught]. I don't view it as a negative....I think [the time spent] has doubled. The reason why is that we're now providing information about objectives as opposed to when we only provided information as to what was your median percentile in reading. We now provide the information as to whether or not, student by student, whether they have mastered certain objectives. So of course, it's a much more concentrated look than it would have been before. So it's doubled. [17] I'm not aware of any [objectives] being neglected" [4833].

VI. Tested objectives get more attention, a necessary trade-off. (14)

30. YYY. [17] "25 percent. It's a trade-off" [1721].

31. YYY. [16] "If the test were not required, I don't think that anyone would spend an unusual amount of time on any objective. [17] Oh gee, not off the top of my head, no I can't. I guess I am generally trying to say, that test from the state is extremely important to us, and if something else has to become of less importance, then so be it. That is the position that we have been put into" [2331].

32. YYY. [16] "Yes, more than I would like to see them doing, but this is true of the State test or any major test because of the emphasis that is placed on it. But you said would they still do this if the tests were not given, I think the objectives would be taught but they might be taught in a different way....[17] I think we have a tendency to emphasize those objectives which are on the test. I don't think we are able to master all of those objectives that are on the test, there are some that even thought they are on the test, which are not taught, and we would say that we don't expect you to teach everything that's on the CAT, but these are the things that we consider important in our curriculum that we do want you to emphasize, so it's kind of a trade-off" [2821].

33. YYY. [16] "There's no question about that. I think [the amount of time] varies. I think there are probably some teachers out there that let the test just about drive their curriculum. Then there's others that just make sure they incorporate the skills into their instruction but don't let it directly drive it. [17] I don't think there's any question that if something is tested it's going to be taught. And if something is not tested it may or may not be taught. I think some of the things that aren't tested probably aren't emphasized maybe as much as the things that are tested" [3732].

34. YYY. [15] "I guess that was one of the efforts. We do change the curriculum sometimes to match the test. In other words, there are times when there's an objective being measured on an achievement test and it might not have been included in the curriculum and then we may add a focused area or something like that, to align it a bit better. Whether that's good or not, it's done. [16] Definitely. I think more emphasis [is placed] on the local program than the state program simply because of the way we can get data back to people so that they know how to use it. [17] I think that may be true in the sense that sometimes the tests are too specific and the skills are too detailed and then we forget the overall goal or global part of what teaching is all about. But I'm not sure if that's a problem, it probably is" [3821].

35. YYY. [16] "Oh yeah. Definitely. I don't know if that's 10 percent as opposed to 5 percent. I couldn't say whether that really drives instruction, but the fact that the test is required and the test results are public certainly influences general teacher behavior in our district. [17] Writing and problem solving aren't readily available on standardized tests. These areas <u>may</u> be less emphasized. I don't think state tests have that much influence in our district, but what influence there is is negative" [3851].

36. YYY. [16] "I would probably say yes, but not intentionally so. Of course, you know, the [State test] is there, and you've got to teach these elements....We require and we document that they teach more than what is considered the [minimum] material. The teacher may have that tendency, but she or he's not allowed to teach just those items. But yes, I know they teach those items for sure because you know that you're going to be tested on them" [4321].

37. YYY. [15] "The state education agency now has a concern that people don't teach the essential elements, they focus on the essential elements <u>that are tested</u>, which is a narrower subset. [16] With the statewide test, yes, definitely. With our norm referenced test somewhat, but not to the same extent. Yes, I think they do spend more time than they would if the test weren't required. [17] I don't know how to answer that in specific terms, I will give you an example. A teacher from a very upper middle class school, probably the highest scoring school in our district on the minimum competency test, claimed that the principal had said to them at the beginning of the year, 'for this year, just forget about the curriculum and make sure the kids know the [State test] objectives.' I don't know if she exaggerated but I know that there was a lot of pressure on principals to have good scores this past year. Other principals are not as sensitive to that kind of pressure, but that's kind of a worst case scenario. Yeah, but I think that we do leave some things out of the curriculum just because of the [press] of time" [4621].

38. YYY. [16] "I'll give you a two-part answer on that one. For the norm referenced test, no. I do not think they spend an inordinate amount of time teaching to those objectives. I think that with the criterion referenced test, the state mandated test, they perhaps do in some classrooms....There has been criticism that the test has begun to be the curriculum, and it is only minimum skills, and there is a great deal of

criticism of the test for that very reason, because there is so much media emphasis and so much evaluation that is based on that, of districts as a whole, of administrators, you know, just overall, and that is one of the reasons it is being revised. [17] Well, I think if anything is, it is in those classrooms where they have concentrated on just minimum skills, finding the details, and that sort of thing. I think higher order thinking skills certainly have been excluded. There has been a great deal of emphasis, of pressure, that teachers have felt, quite frankly, to be certain that they have taught those objectives, and have done it by the month that the test is given. And so to do that, they simply have made decisions to exclude certain objectives" [4711].

39. YYY. [16] "I can tell you, it's required they spend 15 minutes a day....the 15 minutes is supposed to be test taking skills. It heavily emphasizes the state test. [17] A lot of them. The tests only measure the basic skills, reading, math, language arts, writing. There are lots of other areas of the curriculum that are not included, not measured" [4721].

40. YYY. [16] "...I'm not sure how much teaching of specific objectives is actually going on in the schools with the [State] test which is the main emphasis right now. I really think that more has gone into determining which essential elements need to be covered and making sure that those sections of the curriculum are covered in time for testing" [4732].

41. YYY. [16] "Definitely for the State and to a lesser degree for the norm referenced test. [17] I think there's time left in the curriculum for almost all those other important objectives to be covered, and they are covered. But we do have some evidence that shows when you have a basic skills test as we do statewide that the amount of effort that goes into that does subtract from some of the higher level skills. So there is some shifting away from the higher level skills" [4741].

42. YYDK. [16] "Definitely. (State test?) I think they are not as aware of what they should be doing in order to do that; however, if you look at how the test has been designed to match the curriculum frameworks they should be spending the majority of their time covering the content from which that test was designed, so it's difficult for me to know. Those teachers that have really internalized the framework and have made adjustments in their curriculum are probably those teachers whose classes are doing quite well on the State test, and those who have not perhaps had an opportunity or have not made those adjustments are not going out of their way to spend time on that test. We have no organized district effort right now to improve State scores the way some districts do" [4742].

43. YYY. [16] "Well, I naively think that the teachers aren't teaching the specific items of the test so that there may be a few isolated instances where people just don't have their heads screwed on straight. I think that maybe their emphasis on some of the concepts that are on the test is greater than if the test was not required. [17] Obviously, if there are important things that are not covered on the test, they're probably isn't as much feedback to them in terms of them not doing as good of a job, so they might not give the attention to it because they are not 'held accountable for it'" [4841].

Table 2

Quotations Exemplifying the Behaviorist Instruction and Learning Model

Teaching Machines

"How are these reinforcements to be made contingent upon the desired behavior? There are two considerations here—the gradual elaboration of extremely complex patterns of behavior and the maintenance of the behavior in strength at each stage. The whole process of becoming competent in any field must be divided into a very large number of very small steps, and reinforcement must be contingent upon the accomplishment of each step. This solution to the problem of creating a complex repertoire of behavior also solves the problem of maintaining the behavior in strength....By making each successive step as small as possible, the frequency of reinforcement can be raised to a maximum, while the possibly aversive consequences of being wrong are reduced to a minimum" (Skinner, 1954, p. 94).

"Certain experimental studies of variables in programmed instruction pointedly demonstrate the importance of defined objectives to the effectiveness of the instructional enterprise. Falling in this category is the work of Gagne and his collaborators. As this method has developed, it has emphasized not only the specification of the terminal performance, but the analysis of this performance into entire hierarchies of supporting 'subordinate knowledges,' which of course are also performance objectives.

In this series of studies on various tasks of mathematics, it has been shown that the attainment of each of these 'subordinate' objectives by the learner is an event which makes a highly dependable prediction of the next highest related performance in the hierarchy. If a learner attains the objectives subordinate to a higher objective, his probability of learning the latter has been shown to be very high; if he misses one or more of the subordinate objectives, his probability of learning the higher one drops to near zero" (Skinner, 1965, pp 29-30).

Taxonomy of Educational Objectives

"Our attempt to arrange educational behaviors from simple to complex was based on the idea that a particular simple behavior may become integrated with other equally simple behaviors to form a more complex behavior. Thus our classifications may be said to be in the form where behaviors of type A form one class, behaviors of type AB form another class, while behaviors of type ABC form still another class. If this is the real order from simple to complex, it should be related to an order of difficulty such that problems requiring behavior A alone should be answered correctly more frequently than problems requiring AB" (Bloom, 1956, p.18).

Programmed Instruction

"This chapter includes studies which are relevant to the application of programing principles to reading instruction. The organization of this paper differs from the usual division of reading research into such topics as methods, materials, comprehension, and remediation. Instead, the following topics have been used: sequencing factors, stimulus-response factors, reinforcement factors, mediation effects, individual differences, and program evaluations. This structure corresponds with the paradigm of programmed instruction in which desired overt and covert responses are

defined, stimuli are designed to evoke them, reinforcers are applied as needed, items are arranged in a systematic sequence with provision for individual differences in learning rate, and procedures are modified on the basis of learner performance" (Silberman, 1965, p. 508).

Learning Hierarchies

"The existence of capabilities within the learner that build on each other in the manner described provides the possibility of the planning of sequences of instruction within various content areas. If problem solving is to be done with physical science, then the scientific rules to be applied to the problem must be previously learned; if these rules in turn are to be learned, one must be sure there has been previous acquisition of relevant concepts; and so on. Thus it becomes possible to 'work backward' from any given objective of learning to determine what the prerequisite learnings must be—if necessary, all the way back to chains and simple discriminations. When such an analysis is made, the result is a kind of map of what must be learned. Within this map alternate 'routes' are available for learning, some of which may be best for one learner, some for another. But the map itself must represent all of the essential landmarks; it cannot afford to omit some essential intervening capabilities.

The importance of mapping the sequence of learnings is mainly just this: it enables one to avoid the mistakes that arise from omitting essential steps in the acquisition of knowledge of a content area" (Gagne, 1965, 1970, p.242).

Individually Prescribed Instruction

"IPI is based on a carefully sequenced and detailed listing of 'behaviorally-stated' instructional objectives....Each objective should tell exactly what a pupil should be able to do to exhibit his mastery of a given content and skill. This is typically something that the average student can master in one class period. Objectives involve such action verbs as solve, state, explain, list, describe, etc., rather than general terms such as understand, appreciate, know, and comprehend (p.6).

When a student has completed a prescription, he is tested. The test is corrected immediately, and if he gets a grade of 85 percent or better he moves on to a new prescription assigned by the teacher. If he falls below 85 percent, the teacher offers a series of alternative activities to correct weakness, including special individual tutoring. He is not permitted to advance to a new unit of work until he achieves the 85 percent proficiency rating (p.4).

IPI depends heavily on testing. Four types of tests are required: 'wide-band' placement tests to locate unit and level for each student, pre-tests to measure mastery of specific objectives within each unit, post-tests which are alternate forms of the pre-test to determine end of unit mastery, and curriculum-embedded tests to assess within-unit progress" (*Education U.S.A.*, 1968, pp. 11-12).

Mastery Learning

"We have used the ideas of Gagne (1965) and Bloom (1956) to analyze each unit into its constituent elements. These ranged from specific terms or facts to more complex and abstract ideas, such as concepts and principles. They even included complex processes, such as application of principles and analysis of complex theoretical statements. We have considered these elements as forming a hierarchy of learning tasks.

Given our description of the learning tasks for each unit, we have then constructed brief diagnostic-progress tests to determine which of the unit's tasks the student has or has not mastered and what he must do to complete his unit learning. The term 'Formative Evaluation' has been borrowed from Scriven (1967) to refer to these instruments.

The formative tests are administered at the completion of each learning unit and thus help students pace their learning and put forth the necessary effort at the appropriate time. We find that the appropriate use of the tests helps ensure the thorough mastery of each set of learning tasks before subsequent tasks are started. While the frequency of these progress tests may vary throughout the course, it is likely that more frequent formative testing may be needed for the earlier units of the course than for the later ones since typically the early units are basic and prerequisite for all subsequent units. Where the learning of some units is necessary for the learning of others, the tests should be frequent enough to ensure thorough mastery of the former units" (Bloom, 1971, p.58).

Hierarchically Sequenced Learning Objectives

"Briefly, the strategy is to develop hierarchies of learning objectives such that mastery of objectives lower in the hierarchy (simpler tasks) facilitates learning of higher objectives (more complex tasks), and ability to perform higher-level tasks reliably predicts ability to perform lower-level tasks. This involves a process of task analysis in which specific behavioral components are identified and prerequisites for each of these determined (p. 679; cf. Gagne, 1962, 1968).

The order of objectives within each unit is based on detailed analyses of each task. These analyses are designed to reveal component and prerequisite behaviors for each terminal objective, both as a basis for sequencing the objectives and to provide suggestions for teaching a given objective to children who are experiencing difficulty (p. 682).

In practice, implementation of a mastery curriculum implies that children will be permitted to proceed through the curriculum at varied rates and in various styles, skipping formal instruction altogether in skills or concepts they are able to master in other ways. This demand for individualization, in turn, requires that there be some method of assessing mastery of the various objectives in the curriculum....

In our classrooms, the need for assessment is met through frequent testing and systematic record keeping. A brief test for each objective in the curriculum has been written. These tests directly sample the behavior described in the objective" (Resnick, Wang, and Kaplan, 1973, p. 700).

Criterion-Referenced Measurement

"In the late 1950s and early 1960s, a small but plucky band of educational innovators became entranced with the instructional potential inherent in teaching machines and programmed instruction. By transferring some powerful instructional principles, particularly those including a trial-revision teaching model, from the laboratory to the classroom in the form of a carefully sequenced or programmed instruction, these individuals began to achieve startling educational successes. These programmed instruction devotees would start off by explicitly defining a desired post-instruction learner behavior, build a programmed instruction sequence designed to

promote learner acquisition of the behavior, then instruct and posttest learners. If, in rare instances, the instruction proved sufficiently effective in its early form—yummy. But if, as was usually the case, early instructional efforts proved deficient, then the teaching sequence was revised and tried out again with new learners. Because programmed instructional sequences were essentially replicable—that is, were presented to learners by textbook or an audiovisual device in an identical fashion—such trial-revision strategy proved quite effective. Indeed, after a number of revisions it was quite common to secure the kind of shift in performance displayed in Figure 1-3 (a negatively skewed distribution) in which we can see that after effective instruction, the omnipresent normal curve has been bent way out of shape. After truly high-quality instruction, we find few inferior or middling performances—most learners win" (Popham, 1978, pp.12-13).

Appendix B

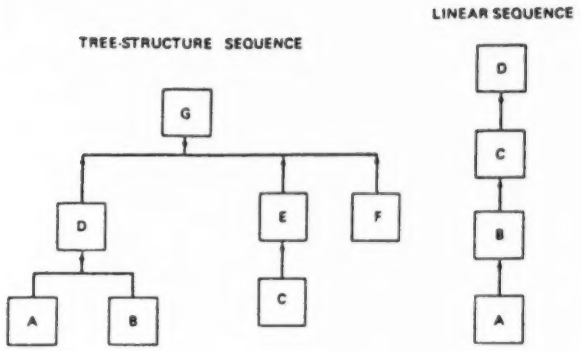

Figure 1: Two possible hierarchies of sequence of instruction from Glaser and Nitko (1971).

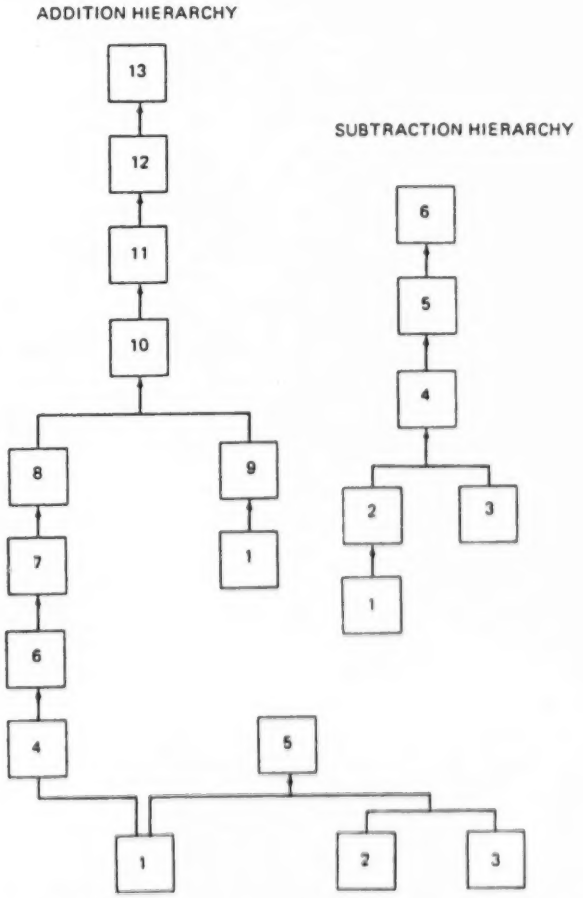

Figure 2: Hierarchies of objectives for an arithmetic unit in addition and subtraction. (Adapted from Ferguson (1969) by Glaser and Nitko (1971)).

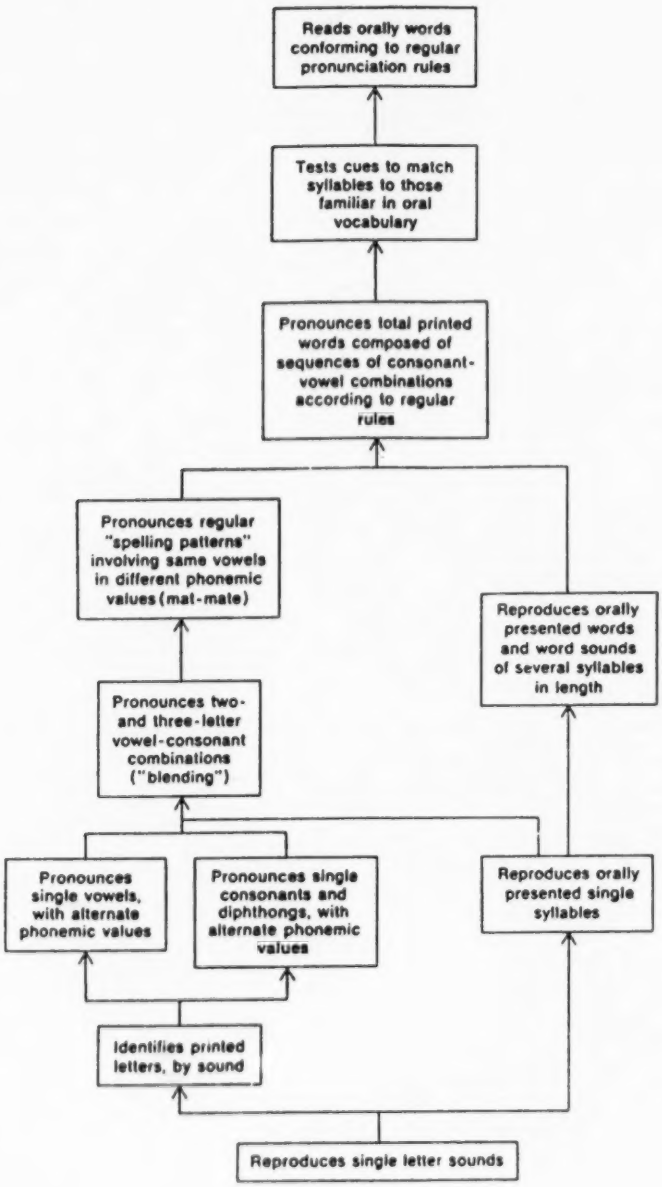
Reads orally words conforming to regular pronunciation rules
Tests cues to match syllables to those familiar in oral vocabulary
Pronounces total printed words composed of sequences of consonant-vowel combinations according to regular rules
Pronounces regular "spelling patterns" involving same vowels in different phonemic values (mat-mate)
Reproduces orally presented words and word sounds of several syllables in length
Pronounces two- and three-letter vowel-consonant combinations ("blending")
Pronounces single vowels, with alternate phonemic values
Pronounces single consonants and diphthongs, with alternate phonemic values
Reproduces orally presented single syllables
Identifies printed letters, by sound
Reproduces single letter sounds

Figure 3: A learning hierarchy for a basic reading skill ("decoding"). (Gagne, 1970).

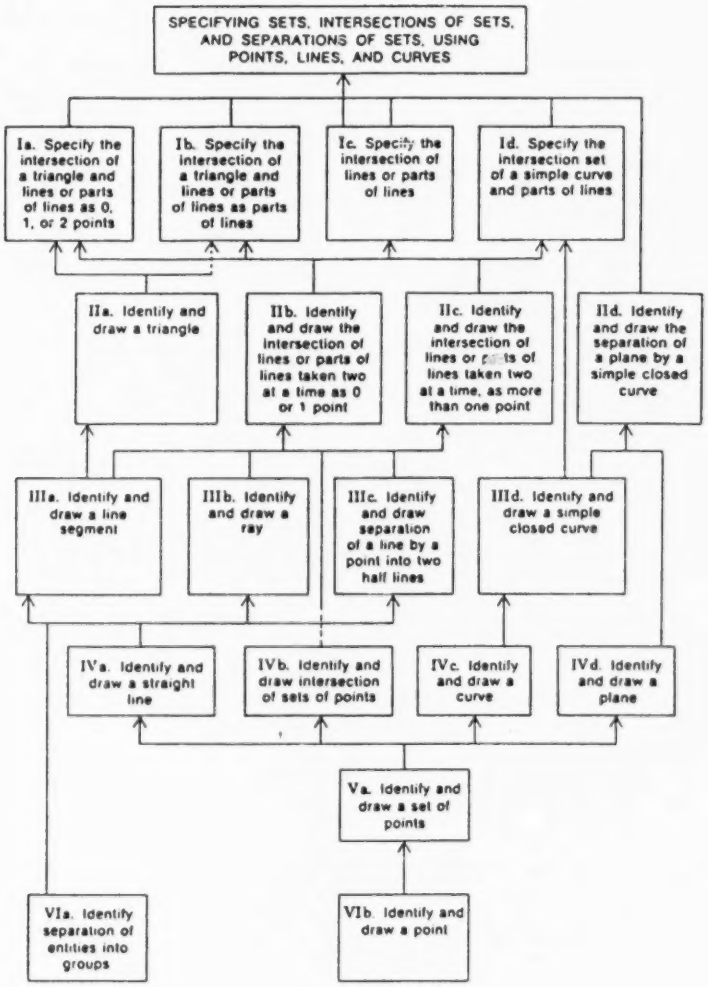
SPECIFYING SETS, INTERSECTIONS OF SETS, AND SEPARATIONS OF SETS, USING POINTS, LINES, AND CURVES
Ia. Specify the intersection of a triangle and lines or parts of lines as 0, 1, or 2 points
Ib. Specify the intersection of a triangle and lines or parts of lines as parts of lines
Ic. Specify the intersection of lines or parts of lines
Id. Specify the intersection set of a simple curve and parts of lines
IIa. Identify and draw a triangle
IIb. Identify and draw the intersection of lines or parts of lines taken two at a time as 0 or 1 point
IIc. Identify and draw the intersection of lines or parts of lines taken two at a time, as more than one point
IId. Identify and draw the separation of a plane by a simple closed curve
IIIa. Identify and draw a line segment
IIIb. Identify and draw a ray
IIIc. Identify and draw separation of a line by a point into two half lines
IIId. Identify and draw a simple closed curve
IVa. Identify and draw a straight line
IVb. Identify and draw intersection of sets of points
IVc. Identify and draw a curve
IVd. Identify and draw a plane
Va. Identify and draw a set of points
VIa. Identify separation of entities into groups
VIb. Identify and draw a point

Figure 4: A learning hierarchy composed mainly of rules (defined concepts) to be acquired in a topic of elementary nonmetric geometry. The topic to be learned is shown in the top-most box. (From Gagne & Bassler, 1963).

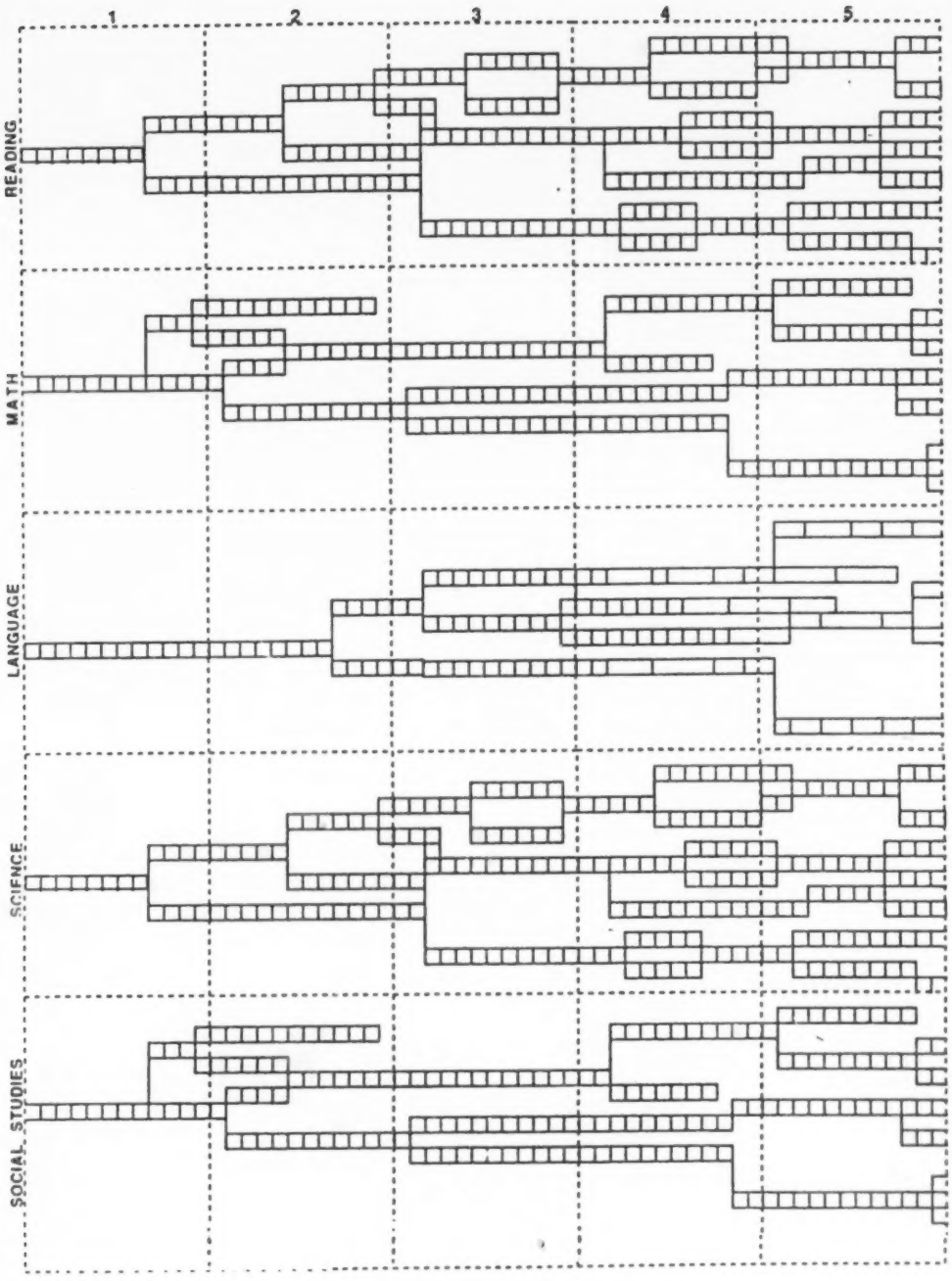

Figure 5: Hypothetical example of parallel sequences of hierarchical objectives.

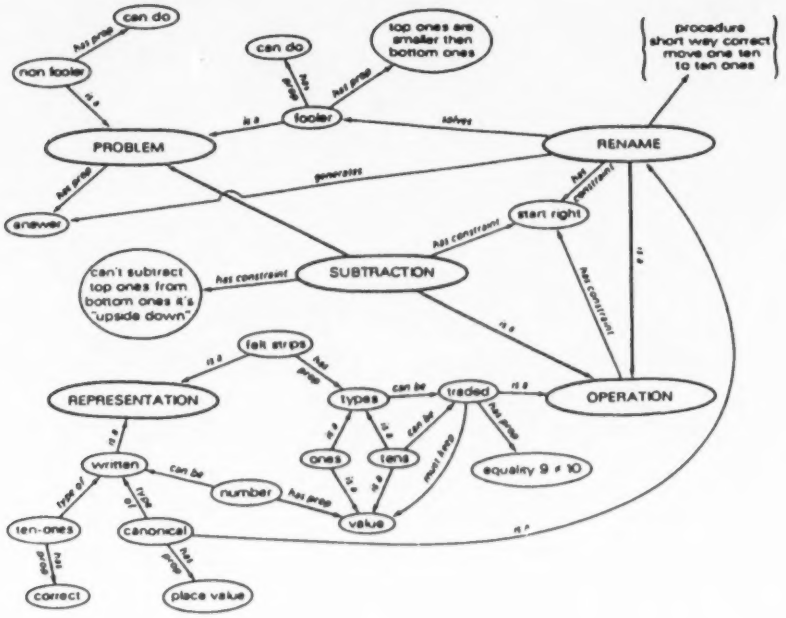

Figure 6: A semantic net representing one child's knowledge after a lesson on two-digit subtraction with regrouping (Leinhardt, 1989).

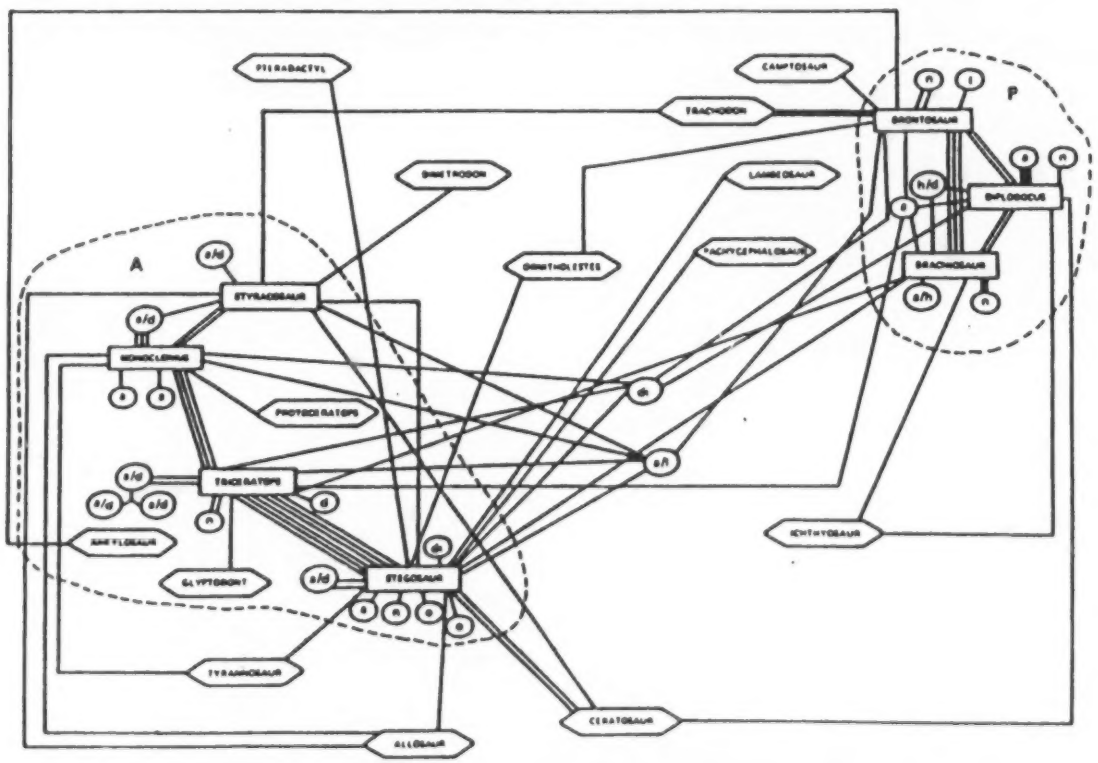

Figure 7: A semantic network representation for better-known dinosaurs for a 4 1/2 year old "expert" (Chi & Koeske, 1983).

(A = armored; P = giant plant eaters; a = appearance; d = defense mechanism; di = diet; h = habitat; l = locomotion; n = nickname; o = other.)

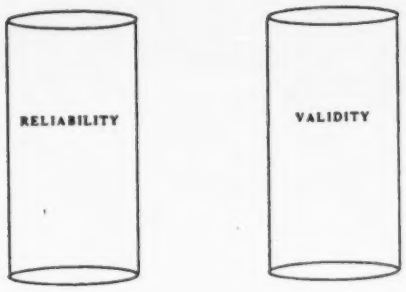

Figure 8: The "twin pillars of reliability and validity in novice conceptions of measurement.

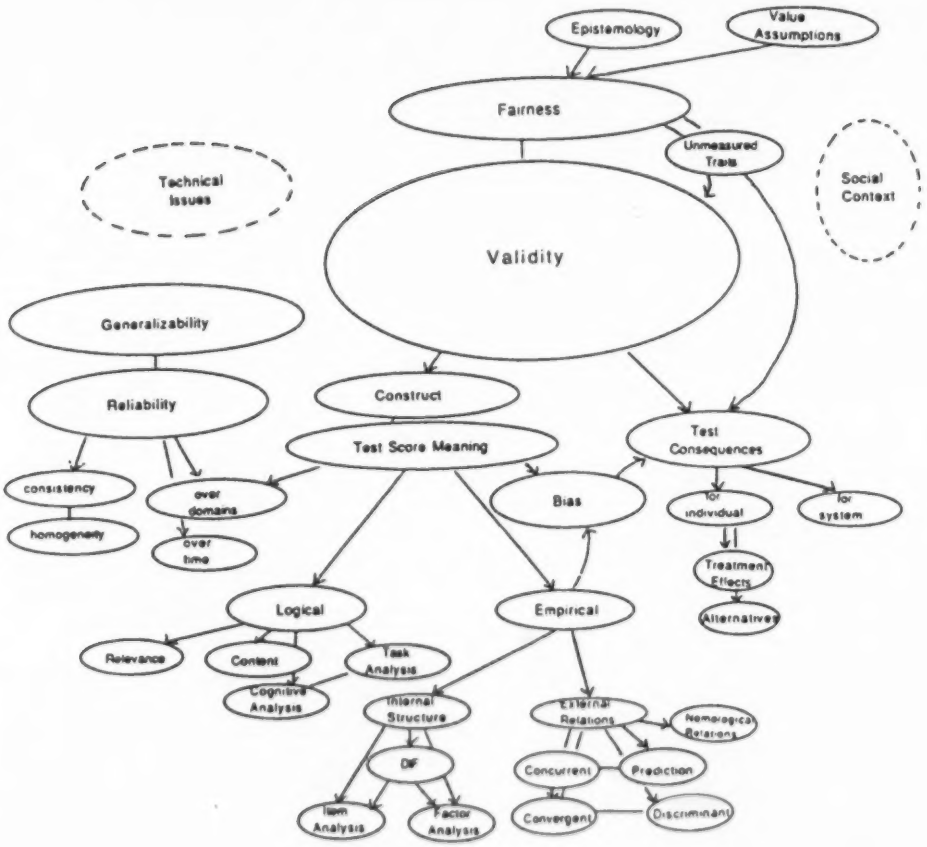

Figure 9: An expert mental model of the measurement concepts associated with reliability and validity.